Evaluating Donor-Level Results Measurement Systems
EPAR Request No. 300

Pierre Biscaye, Melissa LaFayette, Andrew Martin
Monica Richardson, Zoë True

C. Leigh Anderson

*Prepared for the Development Policy and Finance Team
of the Bill & Melinda Gates Foundation*
August 21, 2015

Professor C. Leigh Anderson, Principal Investigator

Highlights of Findings on Donor Results Measurement Systems
- Fifteen of 22 donors describe processes to review and evaluate performance aggregated across all projects, countries, and portfolios.
- The UNG, IADB, Global Fund, DFID, and WBG appear most involved with grantees' monitoring and evaluation processes, to ensure that results data can be aggregated at the donor level. These organizations require a version of a logic model, provide guidance for data standards, and provide data quality assurance for their grantees.
- External reviewers of donors identify aid coordination and alignment between donors and recipients as a priority results area, but few donor reports highlight their progress in aid coordination.
- While most donors appear to be improving their performance in terms of the OECD's indicators of aid effectiveness, the majority of donors still fail to meet many of their individual targets, set by the OECD.
- Multilateral donors specify output and outcome indicators and ensure that data are collected (either by the recipient country or by their own agents) based on those indicators more often than bilateral donors. Both bilateral and multilateral donors are more likely to assess performance in terms of outputs than in terms of outcomes.
- Donors face several challenges for measuring output and outcome results, including ensuring data quality, lack of baseline data, and a focus by both donors and recipient countries on outputs or processes over outcomes.
- Many donor organizations present results in terms of amount spent on various programs. Few donors use cost, output, and outcome data to measure efficiency or cost-effectiveness.
- The majority of donor organizations have rules for reporting donor-level results. Multilateral organizations place a greater emphasis on developing systems for sharing results information.
- Most donors are rated "medium" or "high" for use of results information for strategy and program implementation, but just under a third are rated "medium" or "high" for use of results to influence funding allocations.
- Multilateral donors' attention to outputs and outcomes stands out compared to what is reported by bilateral donors, and multilateral donors are also more likely to compare costs to outcomes in order to evaluate cost-effectiveness
- Differences between the regions where donor organizations are based are slight. Scandinavian donors stand out for their harmonization with recipient country systems and donors from the Americas give the most attention to outcomes and impacts.
- Eight organizations (DFID, GIZ, ADB, GAVI, Global Fund, IADB, WBG, WFP) have strong attention to multiple measures of results, with "medium-high" or "high" ratings in three or more of the following categories of results: coordination with recipients, outputs, outcomes, spending/costs, efficiency, and cost-effectiveness. These eight donors include two bilateral and six multilateral donors.

Abstract

This report is intended to support the Development Policy and Finance (DPAF) team's work on improving the results information that informs aid allocations. Aid results information is often not comparable, since monitoring and evaluation frameworks, information gathering processes, and definitions of "results" differ across donors and governments. This report reviews approaches to results measurement used by multilateral and bilateral donor organizations and highlights trends and gaps in how donors measure and report on their performance. Our review assesses several aspects of donor organizations'

EPAR's innovative student-faculty team model is the first University of Washington partnership to provide rigorous, applied research and analysis to the Bill and Melinda Gates Foundation. Established in 2008, the EPAR model has since been emulated by other UW Schools and programs to further support the foundation and enhance student learning.

results measurement systems, including their institutional design and levels of evaluation for donor results measurement, their organizational processes for measuring different types of results, and their processes for reporting and using results information. We collect evidence on 12 bilateral organizations and 10 multilateral organizations. The evidence we review includes multi-country reviews of aid coordination, peer reviews by other donor organizations, donor evaluation plans and frameworks, and donor results and reporting documents. The report is based on an accompanying spreadsheet that contains the coded information from the 22 donor organizations. Our key findings are summarized in the box above. We find that donors report several types of results, but that there are challenges aggregating certain results across donors, arising from funding and coordination limits for results measurement at the project, country, portfolio, and donor levels. Approaches to results measurement vary across donor organizations. We identify some trends and differences among groups of donors, notably between bilateral and multilateral donors, but overall there are no clear patterns in how donors approach results measurement.

1. Introduction

Bilateral and multilateral aid organizations are interested in understanding whether the funds they provide to recipient countries are used efficiently and achieve desired outcomes. Such information is useful to organizations in learning which of their strategies and programs are effective, identifying elements of programs that are associated with better results, and making decisions about allocating further funding. It is not just donor organizations, however, that have an interest in results measurement, as "A wide range of national and international initiatives and multiple stakeholders call on governments to be more results focused and transparent" (OECD, 2013c). The Millennium Development Goals represent key development results targeted by the international community, while tax payers in donor countries as well as citizens and civil society organizations in developing countries increasingly demand that their governments demonstrate the results of aid spending (*ibid.*).

Efforts to measure results vary, however, according to the purpose and audience for the results information. Aid results information can be used for accountability to external stakeholders, institutional accountability and strategic management, and management and improvement of programs, projects, and policy (*ibid.*). *Figure 1* illustrates a hierarchy of performance measurement, from the least intensive measures, such as activities and outputs which can directly be tied to the organization, to more complicated measures, such as outcomes and effectiveness (measuring outcomes against targets or costs) which require monitoring and evaluation over time. These types of measures serve different purposes, with spending information useful for accountability, output information useful to report specific organizational results, and outcome and effectiveness information useful for strategic management and program implementation. Donor organizations generally fund multiple aid recipients, but results can only be compared to the extent that they consistently measure the same types of results. Since available donor results information varies according to donors' degree of attention to different levels and types of performance measurement, we attempt in this review to analyze donors' results measurement systems and evaluate how they compare to one another.

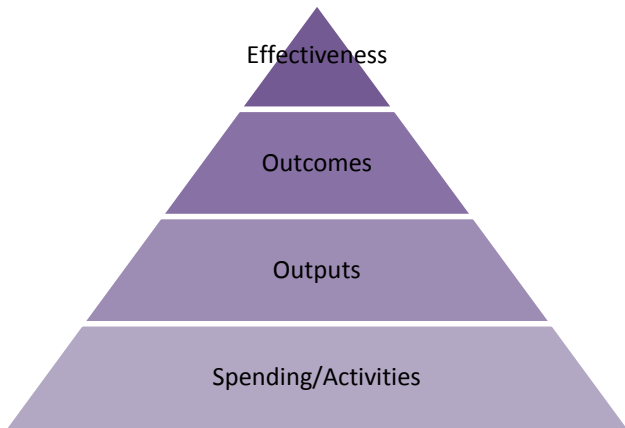**Figure 1: Hierarchy of Performance Measurement**

Figure 1 shows a hierarchy of performance measurement, with the simplest measures, amount of money spent or activities completed, at the bottom and the most challenging measures, outcomes and effectiveness, towards the top. These measures are related and build on each other. For example, measures of outputs can be used to explain changes in particular outcomes, and changes in outcomes are used to evaluate effectiveness. However, donors do not measure higher-level results indicators such as outcomes as consistently as they do spending or activities. This difference is due in part to the greater difficulty of establishing causality in measuring outcomes, but may also relate to a focus on using results information for accountability rather than strategic management.

Agreements such as the Paris Declaration on Aid Effectiveness in 2005 and the subsequent Accra Agenda for Action in 2008 and Busan High Level Forum on Aid Effectiveness in 2011 tie governments and aid organizations together through commitments to improve aid effectiveness. These commitments are intended to create better alignment and coordination

between aid donors and recipients by aligning aid to recipient-country priorities, increasing use of recipient-country monitoring and evaluation (M&E) systems to measure results, and improving mutual accountability. Donor organizations are encouraged to provide technical support to recipient governments and to work closely with them to develop M&E systems that reliably collect and analyze data on aid effectiveness (OECD, 2011). These agreements also bind donor organizations to report a similar set of measures of aid coordination, to better allow for comparisons among donors.

While these agreements outline hypothesized ways to improve aid effectiveness through better coordination, they provide little guidance to donors on how to evaluate effectiveness in terms of outputs, outcomes, or costs. Concepts like results-based management (Binnendijk, 2000) and value for money (DFID, 2011c) are becoming more common among donor organizations, and international targets like the Millennium Development Goals help to focus attention on particular results. In 2010 the Organisation for Economic Cooperation and Development (OECD)'s Development Assistance Committee (DAC) published a set of quality standards for development evaluation, primarily for use by evaluation managers and practitioners at large bilateral and multilateral donor organizations. OECD DAC members are assessed against these standards in periods DAC Peer Reviews (OECD, 2010). A summary of the OECD DAC standards for evaluation is included in Appendix D. However, a 2014 OECD report finds that "progress in meeting international commitments is uneven" and DAC members have faced many challenges in measuring and using results information (OECD, 2014b).

As a result of differences in the purposes and audiences for their results information and in the challenges they face in measuring results, donor organizations have established varying processes for measuring results. We distinguish between four main types of "results" of aid funding that are frequently measured by donor organizations. First, we consider coordination and alignment of aid mechanisms between donors and recipients. This type of "result" is not in our hierarchy of performance measurement, but we include it in this review because of the OECD's emphasis on aid alignment and coordination as part of the recent agreements to promote aid effectiveness. Second, we look at quantities of outputs or services provided as a result of aid funds, the second level of our performance measurement hierarchy. Third, we review measures of changes in outcomes that are related to those outputs or services, the next level up our hierarchy. Finally, we consider measures of spending and of the cost of providing outputs or achieving certain levels of impact on target outcomes. Under this last type of "result" we group everything related to measuring costs, which includes simple measures of spending as well as more complex assessments of efficiency and cost-effectiveness.

In addition to the types of results information that donors measure, this report also reviews how donor results measurement systems are organized, the levels of results that donors report, and donor processes for reporting and using results information. We review peer evaluations and donor documentation describing 22 donor evaluation and results measurements systems, including systems for 12 bilateral organizations and 10 multilateral organizations. Section 2 of this report describes our literature search process and results, and Section 3 outlines our review methodology. In Section 4 we divide our key findings into seven categories, each guided by an overarching question on donor results measurement:

1. Institutional Design: How are donors' results measurement and evaluation systems organized?
2. Levels of Results Measurement: What are the processes for measuring results at different levels of the organization, from the grantee or project level, to the portfolio level, to the donor level?
3. Results - Donor and Recipient Alignment: How do donors measure the Paris Declaration indicators of aid effectiveness, focusing on coordination and alignment between donors and recipients?
4. Results - Outputs and Implementation: How do donors measure activities and outputs, or the implementation of the programs they fund?
5. Results - Outcomes and Impacts: How do donors measure outcomes and impacts of the programs they fund?
6. Results - Costs and Effectiveness: How do donors measure the costs of the programs they fund, and the relative efficiency or cost-effectiveness of different programs?
7. Use of Results Information: What are the processes for reporting results information and for incorporating results in organizational planning, funding, and implementation?

These overarching questions and sub-queries guide the organization of our accompanying review framework spreadsheet and the information we pull from the reviewed documents to populate the spreadsheet. We evaluate general trends from the evidence across all donor results measurement systems and highlight good practices and recurring challenges. In Section 5 we present a summary of different approaches to and definitions of "results" by multilateral and bilateral donors and by donors based in different regions, along with other trends in results measurement. Finally, in Section 6 we summarize our main findings and the principal trends and challenges in aid results measurement. We also discuss gaps in the evidence and potential avenues for future research.

2. Literature Search Process and Results

Our search aimed to identify documents that describe or review the measurement systems of major international aid donor organizations. We conducted one general Google search for information on donor results measurement systems[1], with the understanding that much of the literature we were looking for would be" grey," or unpublished, including reports and evaluations of donors and non-profit organizations. This search captured multiple types of documents referring to donor results measurement systems, including annual reports, reviews, and evaluations of major bilateral and multilateral donors.

In order to ensure that we did not miss any critical published literature, we also conducted searches with the same search string in the Scopus and PAIS International databases. We reviewed all search results but found few relevant documents, suggesting that there is little published information examining donor organizations' results measurement. Appendix A provides more detail on the literature search process.

We also performed individual searches for specific donor organizations' results measurement systems. We targeted the ten largest bilateral donors in terms of 2014 Official Development Assistance (ODA): the United States, the United Kingdom, Germany, France, Japan, Sweden, the Netherlands, Norway, Australia, and Canada (OECD, 2014d). Two other bilateral donors (Austria and Denmark) that appeared in our initial Google search were included in our review. For each of these 12 bilateral donors, we reviewed the results measurement systems of their main aid organizations[2]:

- United States Agency for International Development (USAID)
- United Kingdom Department for International Development (DFID)
- Germany - Deutsche Gesellschaft für Internationale Zusammenarbeit (GIZ – formerly GTZ)
- France - Agence Française de Développement (AFD)
- Japan International Cooperation Agency (JICA)
- Swedish International Development Cooperation Agency (Sida)
- Netherlands - Dutch Ministry of Foreign Affairs (MFA)
- Norwegian Agency for Development Cooperation (Norad)
- Australia - Department of Foreign Affairs and Trade (DFAT – formerly AusAID)
- Canada - Department of Foreign Affairs, Trade and Development (DFATD – formerly CIDA)
- Austria Development Cooperation (ADC)
- Ministry of Foreign Affairs of Denmark (Danida)

Figure 2: Proportion of Total Bilateral ODA from Selected OECD DAC Member Countries, 2013
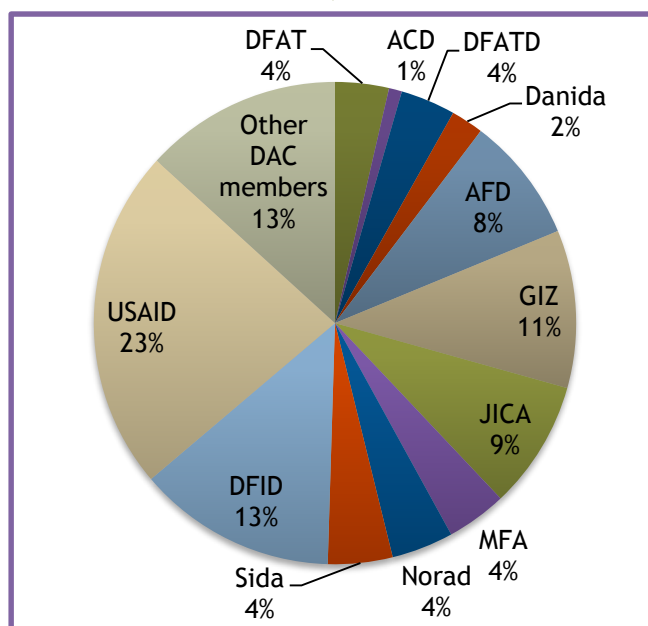
Source: OECD, 2014c.

In addition, we performed targeted searches for seven of the ten largest multilateral donor organizations in 2013 (OECD, 2014c). We did not include the European Bank for Reconstruction and Development (EBRD) or the Arab Funds, and instead reviewed the three larger development banks, the ADB, AfDB, and IADB, and two larger aid funds, the Global Fund and IMF. We also did not include the EU Institutions, as this is a grouping of several smaller organizations with limited documentation describing the results measurement system of the grouping as a single unit. In addition, we also review the GAVI Alliance,

---

[1] We used the following search string: *(system OR framework) AND ("economic development" OR "international development" OR aid) AND ("monitoring and evaluation" OR "performance management" OR "performance monitoring" OR "results management" OR "results data" OR "value for money")*. This search string was designed to limit search results to results measurement systems relevant to international aid and to capture multiple types of documents referring to systems designed to monitor and evaluate results.

[2] These organizations are the primary aid agencies identified on these government's websites, and manage the bulk of ODA from these donors. However, most countries also have other agencies that are responsible for a portion of their international aid.

the 11th largest multilateral donor, and two agencies of the United Nations Group[3]. The ten multilateral donor organizations we reviewed are:

- African Development Bank (AfDB)
- Asian Development Bank (ADB)
- GAVI Alliance (GAVI)
- Global Fund
- Inter-American Development Bank (IADB)
- International Monetary Fund (IMF)
- United Nations Group (UNG)
- World Health Organization (WHO)
- World Bank Group (WBG)
- World Food Programme (WFP)

Figure 2 displays the proportions of bilateral ODA provided by members of the Organization for Economic Cooperation and Development's (OECD) Development Assistance Committee (DAC) in 2013. The 12 donor countries we have selected for review represent 87 percent of total DAC ODA in 2013 (USD 134.8 billion). *Figure 3* presents the proportion of multilateral aid flows provided by the ten multilateral organizations we selected. These ten organizations represent 75 percent of total multilateral aid[4]. All aid totals are those presented by the OECD (2014c) in their statistics on resource flows to developing countries.

Figure 3: Proportion of Total MultilateralAid Flows from Selected Multilateral Organizations, 2013
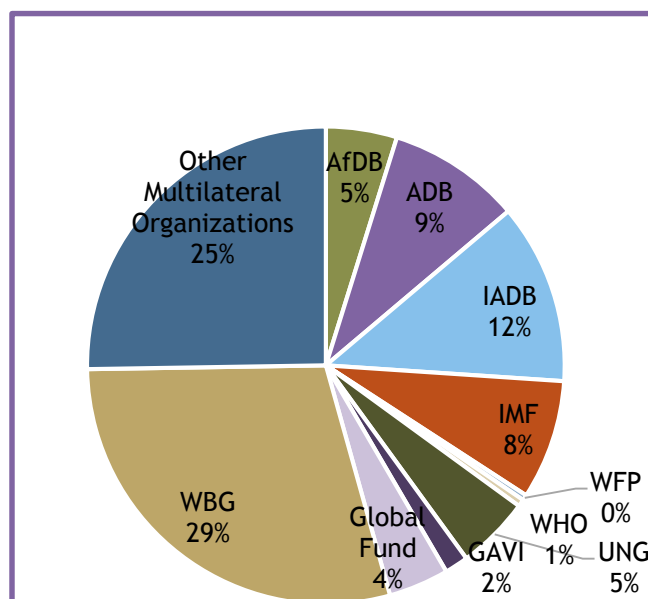
Source: OECD, 2014c.

We searched the websites of each targeted donor organization with the same search string we used on Google. In addition to these searches, we browsed the websites' publications, documents, reports, results, or evaluation pages to identify any additional relevant documents. These searches targeted reports that demonstrated the types of results that donors measure and report, as well as monitoring or evaluation plans describing processes for measuring results. We also conducted targeted Google searches for peer reviews of all selected donor organizations.

For each of our searches, we screened the titles and summaries of the search results and retrieved all documents that met the following screening criteria:

- Reports on the results or performance measurement of at least one specific donor organization;
- Reports on organizational performance OR the system and processes for measuring performance;
- Reports on measurement of aggregate organizational-level performance (as opposed to project-, country-, or portfolio-level performance); and
- Availability in full-text in English.

Our searches yielded 234 relevant documents, including donors' annual reports, evaluation plans, policies for monitoring, evaluation, and performance measurement, and internal evaluations of aid performance. We also retrieved peer reviews of many donors, mostly facilitated by the OECD's Development Assistance Committee (DAC), and several documents reporting on the performance of multiple donor organizations. The evidence we gathered does include information on donor results measurement processes at the project, country, and portfolio levels, but we excluded documents that focused only on these sub-levels to narrow the scope of the review and to emphasize organizational-level results measurement.

Appendix B includes a table summarizing the body of evidence for each of the 22 donor results measurement systems we reviewed. We rated the strength of the body of evidence for each system based on the relevance, quality, and source of the information included in the documents. A "high" rating (13 of 22 donors) indicates that both a recent internal document reporting results information (e.g. an annual report/review) and a recent external review of the organization's evaluation system is available. Bodies of evidence we rate "medium" (seven of 22 donors) are missing either a recent

---

[3] The World Food Programme (WFP) and World Health Organization (WHO) are United Nations agencies. Our coding of United Nations Group is based on system-wide characteristics, but we considered evidence on the WHO and WFP separately, in order to assess whether there are differences in results measurement for the more specialized agencies.
[4] Multilateral aid flows are calculated as the sum of concessional and non-concessional by multilateral organizations is US$ millions.

internal or an external source of information. We rate two bodies of evidence with little recent documentation "low" (MFA, ADC).

3. Review Methodology

Prior to reviewing the selected documents, we developed a review framework covering different components of donor results measurement systems[5]. The review framework questions are derived from a review of the literature on development evaluation. We especially draw on the OECD's quality standards for development evaluation (OECD, 2010), their development results note (OECD, 2013c), their review of measuring and managing results in development cooperation (OECD, 2014b), and their review of results-based management in the development cooperation agencies (Binnendijk, 2000). Appendix D lists the OECD evaluation standards and notes how they are addressed in this report.

We organized our review framework questions (78 in total, listed in Appendix C) into the following seven categories which we discuss in Section 4:
- Institutional Design (11 questions)
- Level of Evaluation (10 questions)
- Results: Coordination/Alignment (20 questions)
- Results: Outputs/Implementation (8 questions)
- Results: Outcomes/Impact (10 questions)
- Results: Costs/Effectiveness (6 questions)
- Reporting and Use of Results Information (13 questions)

To answer the review framework questions for each donor, we reviewed all of the documents describing aspects of that system in order to aggregate relevant information and provide an accurate portrait of each donor's results measurement system[6]. After reviewing all the documents, we summarized the information for each question into the review framework spreadsheet[7].

In the spreadsheet, the majority of questions are coded as either "yes," "somewhat," "planned," or "not specified." "Yes" means there is clear evidence in the documents reviewed. "Somewhat" means the evidence indicates there are some limitations to the results measurement system in a particular area. "Planned" means that internal organizational documents discuss this aspect of results measurement, but there is no evidence of it from organizational reports or external reviews. Finally, "not specified" means that there is no information in the documents reviewed that addresses the question[8]. In some cases, we coded "mixed evidence," which indicates that there is contradicting evidence either within individual documents or across documents, and it was not possible to determine which source was likely to be more valid.

For certain questions about results measurement, we rate donor organizations on a scale from low to high, but the ratings criteria are specific to the individual questions. We describe the criteria used for these ratings as we present our findings. We also use this rating system to describe overall findings for attention to particular areas of results measurement: coordination with recipients, outputs, outcomes, spending/costs, efficiency, and cost-effectiveness.

The review framework spreadsheet is included as a separate deliverable for this project. This report highlights our main findings, which summarize the evidence from multiple documents for each donor organization. When we report information about particular donor organizations, as in lists of donor organizations with particular characteristics, we are citing the entire body of evidence that we reviewed for those organizations. We summarize the characteristics of these bodies of evidence in the review framework spreadsheet, and in Appendix B of this report. We do, however, occasionally cite specific documents to provide examples or more detailed descriptions to support our findings.

---

[5] Our review framework questions and coding are contained in an Excel spreadsheet that aggregates the information from the 22 donor results measurement systems. The spreadsheet also includes short qualitative descriptions justifying the coding decisions for each of the review questions.

[6] For documents specifically describing the particular results measurement system, we read the entire text, while for documents reporting on multiple different systems, we used the search function to locate information about the particular donor organization we were reviewing, or drew information from relevant tables.

[7] The information from the different documents for each country is also available in a single notesheet, organized by question.

[8] The proportion of results measurement systems coded as "Not specified" indicates gaps in our body of evidence, where we lack information to be able to describe different aspects of systems. We did not code "No," as it is possible that our literature search did not identify all documents describing a particular system, and some of the documents we reviewed are several years old, meaning there may have been changes to the systems in the last few years.

4. Findings

4.1. Institutional Design

Mechanisms for creating clear monitoring and evaluation expectations and providing technical support for partners and grantees can influence donor organizations' ability to aggregate and summarize results information from the project or grantee level to the donor level. Institutional design describes how donors organize the process of collecting, aggregating, evaluating, and reviewing results information. We also consider the degree of clarity and guidance donors provide and at what level(s) this is provided (i.e. at the grantee/partner level, recipient-country level, or portfolio level). This section reports our findings in these areas and concludes by discussing the challenges donors face and recommendations donors identified for improvement to these systems. The requirements that donors place on their partners and grantees affect what and how results are reported and ease of aggregating results information across countries or programs. These findings reveal some sources of variation in evaluations of organizational performance at the donor level.

Key Findings on Institutional Design

- Most donors have established results measurement and evaluation processes, and conduct organization-wide evaluations every one to three years.
- Five donors (UNG, IADB, Global Fund, DFID, and WBG) provide grantees with all three types of evaluation support identified by this review: requiring a version of a logic model, providing guidance for data standards, and providing data quality assurance.
- Most donors provide some guidance to grantees on evaluation expectations and standards, however, challenges around reporting remain.

4.1.1. How Donors Evaluate Themselves

Evaluations of organizational performance provide key information for results measurement. The frequency of evaluations and processes used to conduct them determine the use of results information by donor organizations and the immediacy of feedback loops from their grantees.

Figure 4: Frequency of Donor Level Evaluations

- annual
- every 2 years
- every 3 years
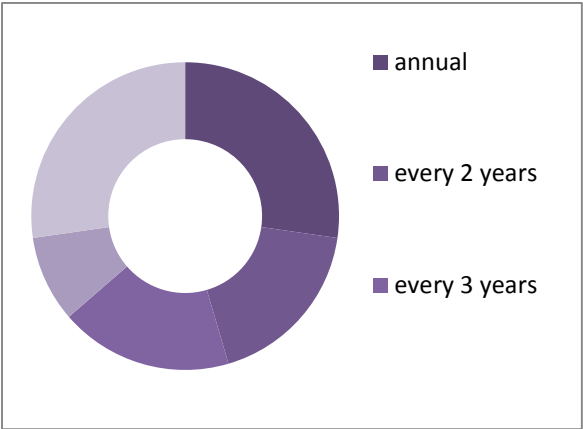
Donor organizations commonly conduct periodic evaluations of their own overall aid performance[9]. Of the 22 donors reviewed, 18 have strategic documents that describe plans to periodically evaluate their own performance. Two-thirds of donors (14 out of 22) conduct evaluations internally, meaning that the donor organization itself carries out the organization-wide evaluation. For five of these 14 donors (WFP, ADB, IADB, AFD, AfDB, Global Fund), the evaluation team is part of an autonomous internal committee that is explicitly separated from the other aspects of the donor organization in order to reduce risks of bias. Eight of the 14 donors that conduct evaluations internally also commission some evaluations from external organizations. Eight donors do not assign evaluation responsibilities to internal groups but rather outsource evaluation activities to independent external reviewers.

16 of the 22 donor organizations report how frequently they evaluate their own performance, as shown in *Figure 4*. Of these, six conduct evaluations annually[10], four conduct evaluations every two years, and four conduct evaluations every three years. Two donors conduct evaluations every five years.

4.1.2. Challenges and Recommendations for Organizational Evaluation

Many donors mention challenges in managing organizational evaluation. **Common challenges relate to lack of financial resources, human capacity, or coordination**. Some donors describe a lack of technical support (GAVI), while others note that staff do not put all of their training into practice due to time and resource limitations (Norad). The WBG describes

---

[9] We distinguish evaluations of donors' overall aid performance, which are intended to assess performance across all organizational programs, from evaluations of individual programs, projects, or portfolios, which are generally conducted more frequently.
[10] One, Sida, plans to conduct annual evaluations, but we could not confirm that they have begun doing so.

more specific challenges around technical support: "World Bank Group staff sometimes face incentives that constrain the scope and relevance of IE(s) [impact evaluations]. This includes incentives to assess interventions that are "easier" to evaluate, limited understanding of the tool, or fear of negative results." (IEG, 2012a) Over 40 percent of survey respondents indicated that lack of staff capacity was a constraint for conducting impact evaluations in the WBG (*ibid.*). The Asian Development Bank describes coordination issues between the grantee staff and the Bank's Independent Evaluation Department (ADB, 2014c). Funding for evaluation activities is a challenge for several donors. A 2013 USAID report postulates that budget reductions in 2009 caused evaluation contracts to be awarded to the lowest bidder, and therefore led to lower-quality evaluations. Reports from the WFP and AfBD indicate that these organizations are unable to sufficiently fund evaluation and oversight positions (UNEG, 2014; AfDB, 2013d).

Recommendations for addressing the capacity for evaluations largely centers on staffing, training, and time allocation. For staffing and training, recommendations include creating a comprehensive training for program staff (Norad), more clarity on what should be measured (Norad), selecting staff to receive intensive evaluation and results training (AFD), or having an evaluation specialist on each evaluation team (USAID). Time allocation recommendations include budgeting for and planning evaluation time into program management decisions (DFAT), and setting a cut-off time for reporting to allow for distinct periods for evaluation (ADB).

4.1.3. How Donors Organize Results Measurement with Grantees

Donor organizations often take responsibility for providing clear expectations and guidance to grantees and partners on their preferred approaches to results measurement and evaluations. These preferences may include being able to attribute impacts, and some donors devote considerable resources to monitoring and evaluation systems in order to be able to attribute change to particular interventions in comparison to what would have happened if the interventions had not occurred. This guidance starts with the intervention design and proposal where a clear pathway between the intervention and the intended impact is drawn out into a theory of change, logic model, results chain, or results framework. Once the intervention is underway, expectations or standards around data collection and aggregation support grantees in reporting results that are useful for the donor organization. Many donors also provide technical support and systems for data verification to promote quality during collection and reporting the results information.

Five organizations (UNG, IADB, Global Fund, DFID, and WBG) appear most involved with grantees' results measurement processes, requiring some version of a logic model, providing guidance for data standards, and providing data quality assurance. Of these five donors, only DFID is a bilateral donor.

Nearly 60 percent of donors (13 out of 22[11]) require their grantees to develop a **theory of change, logic model, results chain, or results framework**. A DFAT report notes internal pressures to formalize guidelines for developing logic models (DFAT, 2014b). A 2014 peer review of Danida reports efforts to invest in institutionalizing a theory of change (Kvalvaag & Stern, 2014). A 2014 overview of evaluation at the IADB mentions that before implementation, all IADB projects are assessed in terms of "evaluability" using a Development Effectiveness Matrix, and notes that one aspect of the "evaluability" of a project is the use of a theory of change and results matrix in the project plan. The review states that "The IDB Development Effectiveness Matrix (DEM) measures the upstream evaluability of all operations, rating the following aspects of each project on a scale of one to 10: the diagnosis of the development problem to be addressed, the logic of the selected intervention, the results matrix for the intervention, the economic analysis of the intervention, and the monitoring and evaluation plan" (IADB, 2014a).

Half of donors (11 of 22) provide grantees with **guidelines or standards for data collection or aggregation**. For example, the AfDB notes that "the Approach Paper sets out the issues, expected focus and outcome, key questions and evaluation methods. It covers data collection methods and sources of information and includes the proposed schedule of activities, resource requirements, and plan for communication and dissemination of findings and recommendations" (IDEV, n.d.). The 2014 IMF Annual Report indicated that "In February 2014, the IMF developed a draft standard template for countries to use for the collection of data on government revenues from natural resources, based on the revenue classifications of the Government Finance Statistics Manual 2001" (IMF, 2014b). Three other donors provide only some[12] guidance (Sida, DFATD, and USAID), and one planned to provide guidance (AFD). According to an evaluation of USAID evaluations, most USAID evaluations (90 percent) described data collection methods, but only 23 percent linked data collection methods to the

---

[11] Two, DFAT and AFD, describe plans to require logic models or results frameworks in the coming years in documents from 2013.
[12] "Some guidance" is coded when the donor provides guidance for some but not all areas (DFATD) or when standards or guidance is incomplete in some way (USAID, Sida)

research questions. The explanation of data transfer to USAID was rated "weak", though the review noted that this was a new standard (Hageboeck, Frumkin, & Monschein, 2013).

Another area where some donors provide oversight to grantees is **data quality assurance and verification**. Ten of 22 donors provide some support in this area, the majority of which also provide guidance to grantees on data collection or aggregation[13]. Methods used to verify data include independent internal committees for data review and grantee-based mechanisms, such as self-assessment or peer assessment of validity. As an exception, Danida used external evaluators to verify their data. Their 2012 Evaluation Guidelines stress the importance of data verification and data quality: "The first step in the analytical process involves the cleaning, verification and organization of data, so that they meet standards of quality, i.e. validity and reliability. Data also need to be organized according to the evaluation criteria and questions" (Danida, 2012b).[14] In some cases, data verification is ensured through technical advice provided by the donor (e.g. Norad). A desk study of Norwegian aid evaluation notes that the "current model, whereby individual programme officers either quality assure results frameworks themselves or seek technical advice from Norad, leads to inconsistencies in the robustness with which proposed results information is being accepted and reported" (Wiig & Holm, 2014). In other cases, general guidance on data verification may be provided, but it is unclear how much it is used (e.g. Sida), while other agencies allow grantees to come up with their own systems (e.g. ADC).

A 2013 review of aid program performance reports (APPRs) of Australia finds that the template used provided little guidance on expectations for data quality (DFAT, 2013). For the WFP, a 2014 peer review finds that "The Evaluation Quality Assurance System (EQAS) has played a major role in ensuring the consistently high quality of OEV evaluations and gives WFP perhaps the most comprehensive evaluation management and quality assurance process in the United Nations system" (OECD, 2014). USAID uses a triangulation method for data verification, with partner organizations as one source. This implies that partner organizations are given guidance, standards, or support to effectively collect data (USAID, 2013e).

4.1.4. Challenges for How Donors Organize Evaluations with Grantees

Though grantee results measurement and reporting is important to donors' ability to assess overall performance, it remains a common challenge. Fewer than half of donors (nine of 22) provide standards or guidelines for reporting. Even when formal methods exist, such as Norad which provides reporting templates, enforcement of these methods is weak (Lloyd, Poate, & Villanger, 2014). Lack of guidance and reporting standards provided to grantees limits organizations' ability to aggregate results information at the donor level. Regular detailed reporting, such as reporting expenditures, budgets, targets, and lessons learned, does not necessarily mitigate the issue of inconsistent reporting standards. For example, a 2000 OECD review of results-based management notes that although DFAT required detailed annual reporting, variability existed in the reports' quality (Binnendijk, 2000). Additionally, reporting can vary depending on the program level and how that interacts with the recipient-country level (GAVI, 2011).

4.2. Levels of Results Measurement

Key Findings on Levels of Results Measurement
- Most donors describe a process for results measurement at more than one level, usually the donor level and one lower level. Only three (Global Fund, DFAT, and DFATD), describe a process for both project-level and country-level results measurement.
- Nearly 70 percent of donors describe a process to review and evaluate aggregate performance at the donor level.
- Access to experts, lack of funding, and capacity are the most cited challenges among donors.

Results measurement is based on data collected from the point of program implementation and aggregated up to evaluate performance at the donor level (*Figure 5*). Organizations provide guidance and support at different levels to promote coordination and alignment of results measurement: at the project level (individual programs), the recipient-country level (collections of all programs within a country), and the portfolio level (collections of particular programs within and across countries). The decision to report findings at different levels depends on the expected uses of results information for each organization, and on the information demands of their stakeholders. While our search process excluded documents that focused exclusively on project-, country-, or portfolio- level results measurement, the evidence we reviewed does include

---

[13] Nine of the ten provide support in both areas. One, USAID provides data verification support, but is only rated "somewhat" for providing guidance on data collection/aggregation.
[14] It appears these guidelines are for the external evaluators; it is not specified whether grantees themselves use these methods.

information on results measurement processes at these sub-levels. This
section reports our findings on the support or guidance provided and the
challenges faced by donor organizations at these different levels of results
measurement.

Figure 5: Hierarchy of Results Measurement
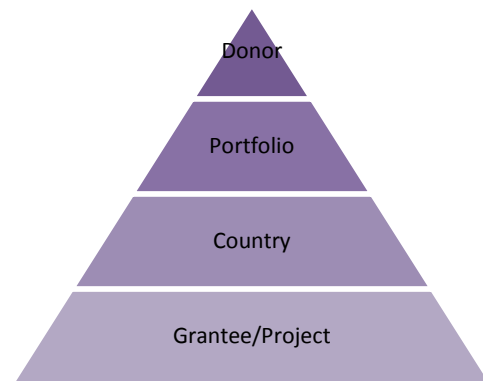
4.2.1. Project-Level Results Measurement

**Most donors (15 of 22) describe processes to monitor and evaluate
results at the project level**, though the specificity varies[15]. AFD conducts
project evaluations to ensure accurate information on resources used,
project achievements, and results (AFD, 2013). A similar process is used
by JICA (JICA, 2014b). Other donors describe providing an element of
oversight through a contact person (USAID) or review by the evaluation
committee who validated the reports (ADB). Some donors, such as the UN
Group, mentions using concrete tools, including checklists (UNEG, 2010).
At the ADB, departments create Project Completion Reports. When the ADB's Independent Evaluation Department wants to
verify the objectivity and validity of grantees' self-evaluations, they check a sample of these reports and prepare an
independent Project Performance Evaluation Report (ADB, 2013c). The WFP has an Evaluation Policy document which
describes standards for project level evaluations. At the WFP, the Office of Evaluation writes centralized evaluations of
projects, while decentralized evaluations are prepared by recipient-country offices with external consultants (UNEG, OECD,
& DAC, 2014). IADB's process for project-level results measurement appears the most complete and includes details on the
review process and which stakeholders are involved (IADB, 2014b)[16].

Challenges for collecting and reviewing project-level results are due largely to inadequate data collection systems. Danida,
the WFP, and the Global Fund report challenges with missing baseline data or appropriately disaggregated data. DFID
expresses concern over balancing the priorities of creating a sense of ownership among the grantees and quality of data
(OECD, 2014a). These challenges in results measurement at the project-level make it more difficult for organizations to
aggregate results for measuring at higher levels. Other challenges include include developing effective standards and
guidance for staff (DFID, Norad, AFD, UNG) and that making clear connections between project goals and strategies and
results to be measured (ADC, Danida)

Recommendations for improving project-level results measurements vary and are largely tied to individual organizations'
specific challenges. Recommendations include doing peer-reviews of economic and financial findings (ADB), developing a
culture of evaluation and methods to better evaluate more challenging interventions (AFD), avoiding vague statements in
reports (USAID), and paying more attention to cost and value (IADB). The objective of these recommendations is to better
manage project-level monitoring systems in order to ensure collection of relevant results data.

4.2.2. Country-Level Results Measurement

Just as many donors have processes for reviewing and collecting results information for individual projects, donor
organizations may also have processes for measuring results at the country level, across multiple programs. **Fifteen out of
the 22 donors report processes for country-level results measurement** in the documents we review, thought only three
(Global Fund, DFAT, and DFATD), describe a process for both project-level and country-level results measurement.

In many cases, donors aggregate and summarize country-level results information in their annual reports. Some donors,
however, have specific evaluation programs for the country level. In Canada, DFATD reviews all countries within each five
year rolling evaluation program (DFATD, 2014a), and the ABD creates Country Assistance Program Evaluations (Asian
Development Bank, 2013a). In Australia, DFAT uses a formal evaluation tool (aid program performance reports) at the
country and regional program levels (Department of Foreign Affairs and Trade, 2013). The Japanese Ministry of Foreign

---

[15] Our review focused on organizational reporting and results measurement and therefore did not focus on project-level results
measurement practices, except to the extent that these were described and reported in organizational reports and protocols for results
measurement.

[16] "All of OVE's [Office of Evaluation and Oversight] evaluations follow a formal process for preparation, review, Board discussion, and
disclosure, which is laid out in various IDB governance and administrative documents. Each begins with an Approach Paper that lays out the
evaluation questions and methodology. Approach Papers are reviewed and discussed with IDB management and then finalized, sent to the
Board, and disclosed to the public. Drafts of final evaluations are reviewed and discussed with IDB management and, in the case of country
program evaluations, with the country authorities. They are then finalized by OVE and sent to the Board for discussion and subsequent
disclosure…" (IADB, 2014).

Affairs conducts evaluations of JICA's performance at the country level with respect to "relevance of policies," "effectiveness of results", and "appropriateness of processes" (JICA, 2014a). The 2014 Annual Development Effectiveness Review of the AfDB reports that they "undertake regular Country Portfolio Performance Reviews to tell us whether our programmes are on track" (AfDB, 2014). The Independent Evaluation team of the AfDB collects results information and provides reports at the country-level as well. Their three-year rolling evaluation plan "includes evaluations of projects and programmes, sector and thematic evaluations, recipient-country and regional strategy evaluations, corporate evaluations, joint evaluations, and capacity development and knowledge management activities for effective feedback, dissemination and learning" (IDEV, n.d.).

Ten donors mention challenges with country-level results measurement. Data quality is an issue for several organizations (ADB, AfDB, Global Fun, WFP), as is lack of availability of baseline data (ADB, DFAT, MFA, WFP). Four donors (ADB, Global Fund, UNG, WBG) describe the importance of a clear results framework in collecting and reviewing results at the recipient-country level. Two other organizations (ADB, DFATD) note the need to align time-trend indicators and ensure that the indicators tracked are harmonized across programs. Three donors (MFA, DFID, WBG) also report issues with the harmonization between country goals and donor strategy.

4.2.3. Portfolio-Level Results Measurement

Eighteen of 22 donors describe a process to review and collect results at the portfolio level. Measuring and aggregating results across multiple programs in specific thematic areas supports organizations in evaluating and summarizing their overall performance. Reporting of results at the portfolio level is therefore common in many organizations' annual reports. For example, Norad's 2012 and 2013 annual reports describe portfolio evaluations for agriculture and food security, climate and forest, East Africa, people with disabilities, and oil for development (Norad, 2013). The Independent Evaluation Department of the ADB conducts sectoral and regional evaluations: "A sector/thematic evaluation study focuses on a selected sector or theme issues across countries or a subregion. A corporate evaluation study evaluates a policy, modality, or process relevant to ADB operations" (ADB, 2013c).

Eight organizations report challenges with or recommendations for portfolio-level results measurement. Challenges described in the documents we reviewed are similar to those at the country level: issues of inadequate baseline evaluations (Norad, DFAT), alignment of indicators (DFATD), lack of results frameworks (Norad), and lack of harmonization of strategies and goals (USAID). Some donors face multiple challenges at once. For example, the 2013 Norad annual report lists several challenges, including a focus on short-term results, weak monitoring and evaluation, lack of baseline data, and absence of results frameworks to guide evaluations (Norad, 2013).

Recommendations for improving results measurement at this level include using gender-specific evaluations (USAID, IADB), providing more guidance for data collection and aggregation (USAID), using an interrupted time-series model when appropriate (Global Fund), developing coherent theories of change (Norad), and considering the lag time for impact (Global Fund).

4.2.4. Donor-Level Results Measurement

Results information from the project, country, and portfolio levels is useful for management and improvement of programs, projects, and policy, and also for institutional accountability and strategic management. Results at these levels are less useful, however, for high-level political accountability and results reporting. For this, organizations aggregate and assess results information at the overall donor level. **Eighteen of 22 donors describe processes to review and evaluate performance aggregated across all projects, countries, and portfolios.** Results at the organizational level are usually presented in annual reports. For example, IMF Annual Reports summarize performance in a variety of areas, including surveillance, financial support, capacity development, risk management, internal governance, staffing, and transparency (IMF, 2015a). The information in the remainder of this report describes result measurement processes at the organization-level.

An additional three organizations (MFA, Sida, UNG) describe plans to institute a process for aggregating results at the donor level, in addition to their current attention to results at lower levels. For example, a 2013 DAC peer review indicates that Sida has been slow to shift evaluations toward strategy and management, despite DAC recommendations to do so (DAC, 2013). In the Netherlands, the MFA is supposed to publish reports every two years, but it has not done so since 2010 (Ministry of Foreign Affairs of the Netherlands, n.d.). The UN Evaluation Group describes plans to support and independent system-wide evaluation, but indicates that this has not yet occurred (United Nations Evaluation Group, 2013a).

Over half of the donors (12 of 22) mention challenges in collecting, reviewing, and aggregating donor-level results. Many challenges are similar to those in other levels of results measurement, but organizations report specific challenges with aggregating results at the donor level, including attribution (DFATD)[17], lack of baseline data and financial data (Gobal Fund), communication across agencies (ADC), lack of a learning culture (DFID), excessive flexibility over monitoring and reporting (Norad), and concern over assessing longer term results (WFP).

Recommendations for donor-level results measurement include: clarify employee incentives and consequences around prioritization of evaluation (Norad), clarify policies for validation of data (ADB), use a logical framework connecting strategies with desired results (AFD), and use real-time monitoring systems (Global Fund).

4.3. Results - Donor and Recipient Alignment

A primary emphasis of the Paris Declaration on Aid Effectiveness of 2005 is "alignment" and coordination of donor and recipient-country monitoring and evaluation systems (OECD, 2011). Donor-country alignment refers to the use of in-country systems for financial management and monitoring and evaluation (instead of using parallel implementation units), as well as making sure aid flows are in line with recipient-country development agendas. When donor and recipient-country monitoring and evaluation systems are not aligned, in-country systems may be overburdened with multiple reporting requirements, grantees may be required to purchase goods from sources in the donor country ("tied" aid), and aid may not be effectively used to assist with a recipient-country's development goals.
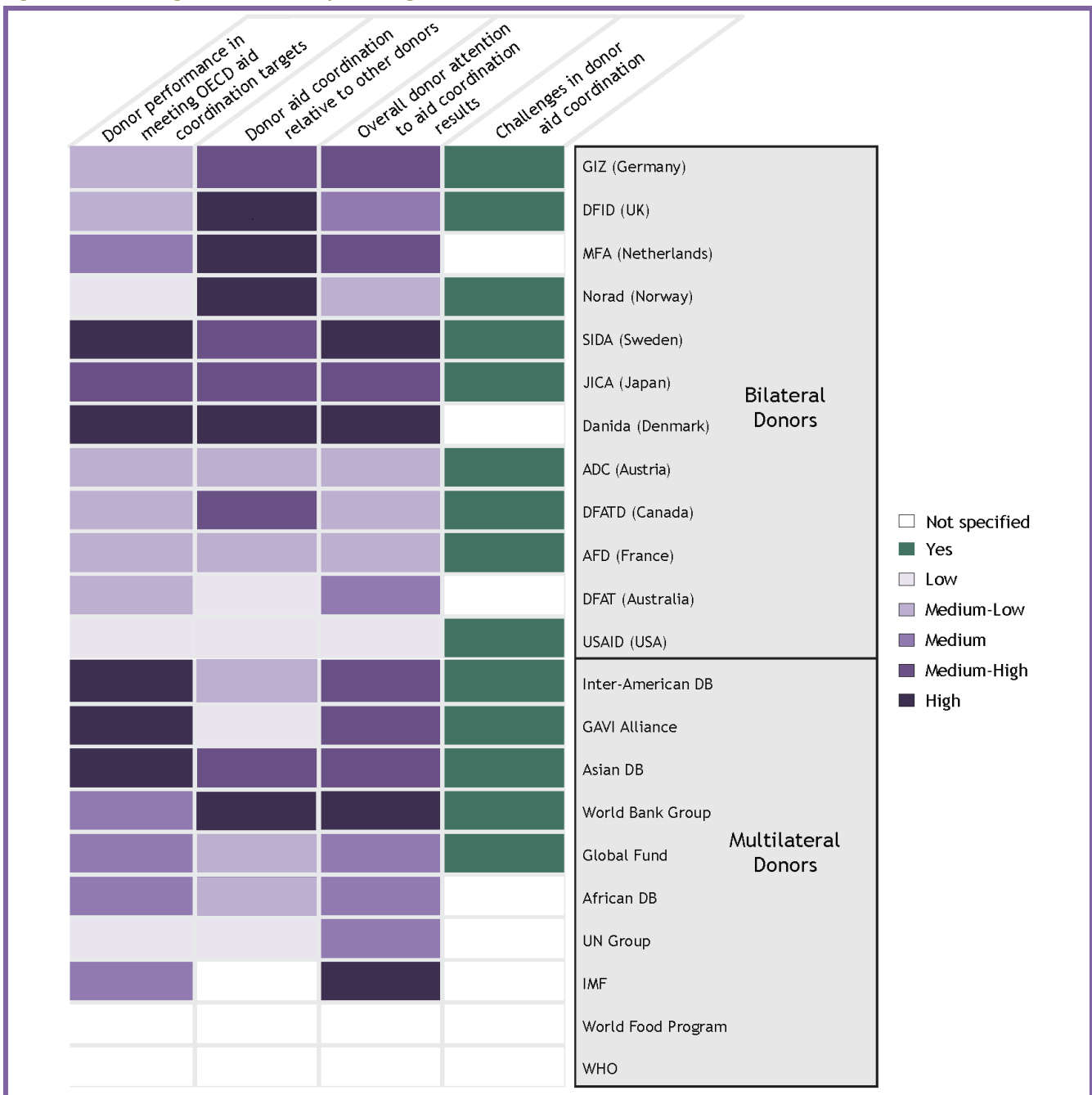
Key Findings on Measuring Donor and Recipient Alignment Results
- The value of donor-recipient coordination is emphasized in many external reviews of donors, especially those organized by the OECD, but fewer donors report on their performance in aligning results measurement.
- While most donors appear to be improving their performance in terms of the OECD's indicators of aid effectiveness, the majority still fail to meet many of their individual targets, set by the OECD.
- Challenges to aid coordination described by donors generally fall into three categories: taking action to improve coordination, coordinating with numerous diverse stakeholders, and coordinating joint evaluations.

As a result of the Paris Declaration and later international agreements on how to support aid effectiveness, indicators of coordination and alignment between donors and recipients have gained importance as measures of donor performance. We therefore begin our review of the types of results measured by donor organizations by assessing their attention to measuring donor and recipient alignment. We consider donors' performance on several measures of alignment as an indicator of their attention to this type of results measurement. In addition, we review common challenges in coordination reported by donor organizations. *Figure 6* is a heat map summarizing our findings on alignment results. In the heat map, donors with similar qualities are clustered together.

---

[17] DFATD documents noted challenges in attributing evaluation findings to implementation of interventions, especially when assessing higher-level indicators.

Figure 6: Measuring Donor and Recipient Alignment Results

	Donor performance in meeting OECD aid coordination targets	Donor aid coordination relative to other donors	Overall donor attention to aid coordination results	Challenges in donor aid coordination	
	Medium-Low	Medium-High	Medium	Yes	GIZ (Germany)
	Medium-Low	High	Medium-Low	Yes	DFID (UK)
	Medium	High	Medium		MFA (Netherlands)
	Low	High	Medium-Low	Yes	Norad (Norway)
	High	Medium	High	Yes	SIDA (Sweden)
	Medium-High	Medium	Medium	Yes	JICA (Japan)
	High	Medium-High	Medium		Danida (Denmark)
	Medium-Low	Medium-Low	Medium-Low	Yes	ADC (Austria)
	Medium-Low	Medium-High	Medium-Low	Yes	DFATD (Canada)
	Medium-Low	Medium-Low	Medium-Low	Yes	AFD (France)
	Medium-Low	Low	Medium		DFAT (Australia)
	Low			Yes	USAID (USA)
	High	Medium-Low	Medium	Yes	Inter-American DB
	High	Low	Medium	Yes	GAVI Alliance
	High	Medium-High	Medium	Yes	Asian DB
	Medium	High	High	Yes	World Bank Group
	Medium	Medium-Low	Medium	Yes	Global Fund
	Medium	Medium-Low	Medium		African DB
	Low		Medium		UN Group
	Medium		High		IMF
					World Food Program
					WHO

Bilateral Donors (GIZ through USAID)

Multilateral Donors (Inter-American DB through WHO)

Legend:
- ☐ Not specified
- ■ Yes
- ☐ Low
- ■ Medium-Low
- ■ Medium
- ■ Medium-High
- ■ High

4.3.1. Indicators of Attention to Coordination

We draw on several sources to assess donor organizations' attention to coordination and alignment with grantees and implementing partners. First, we review a 2011 OECD review of aid effectiveness which compares donors' performance against targets for several indicators of coordination (OECD, 2011). Second, we review a 2010 aid effectiveness index based on these OECD indicators, which compares donors' performance in aid coordination against other donors (Sinha, 2010). Finally, we consider evidence of attention to coordination from donor reports and peer reviews.

The 2011 OECD "Aid Effectiveness 2011: Progress in Implementing the Paris Declaration" review includes an appendix with tables reporting on several indicators of aid effectiveness for each donor[18], with the exception of the WHO and WFP whose results are combined with the other organizations comprising the United Nations Group[19]. Data are drawn from the 2011 Survey on Monitoring the Paris Declaration, and are collected at the country level through government and donor questionnaires. These are the most recent available data on these measures. A total of 78 countries participated in this survey, though not all recipient-country programmed aid was included in this survey[20]. The donor tables also included 2010 targets for each donor[21].

Over half of donors we review meet their targets for coordinating technical assistance (ten of 18) and using recipient-country public financial management systems (11 of 20). Fewer donors, however, meet their targets for number of parallel project implementation units (PIUs) (seven of 20), coordinating donor missions (six of 20), and coordinating recipient-country analyses (three of 20).

We assign donors a rating from low to high based on donors' overall performance in meeting their targets for the following six indicators: strengthening capacity by coordinated support, using recipient-country public financial management systems, avoiding parallel implementation structures, and ensuring untied aid, joint recipient-country missions, and joint analytic work. Our findings are displayed in column one of *Figure 6*. Donors who meet their targets on one or fewer indicators of coordination and alignment are rated "low." Those that meet two targets are rated "medium-low," those that meet three or four targets are rated "medium-high," and those that meet five or more targets are rated "high." According to these criteria, three donors are rated "low" (Norad, UNG, USAID), nine are rated "medium-low" (ADC, AFD, AfDB, DFAT, DFATD, DFID, Global Fund, GIZ, IMF, WBG), five are rated "medium-high" (ADB, GAVI, IADB, JICA, MFA), and two are rated "high" (Danida, Sida) for aid coordination.

We also consider a composite "overall aid effectiveness index" (Sinha, 2010) which ranks donors based on their performance on the 2008 Survey on Monitoring the Paris Declaration, displayed in column two of *Figure 6*. A donor's overall rating on this index is the mean of their rating for each indicator of aid effectiveness, calculated using the formula *indicator rating = (actual donor value – minimum value) / (maximum value – minimum value)*[22]. Thus, for each indicator the highest-performing donor receives a score of 1, while the lowest-performing donor receives a score of 0. Sinha (2010) uses this method to calculate an aid effectiveness index for 31 donors, including all the organizations we review except the WHO, WFP, and IMF. Five of the donors we review (Danida, DFID, MFA, Norad, WBG) have an index score above 0.7 (top seven of the 31 ranked donors). Another five donors (ADB, DFATD, GIZ, JICA, Sida) have a score from 0.6 to 0.69 (next seven donors). Five more donors (ADC, AFD, AfDB, Global Fund, IADB) have a score from 0.5 to 0.59 (next seven donors). Finally, four of the donors we review (DFAT, GAVI, UNG, USAID) have a score below 0.5 (remaining ten of the 31 donors).

4.3.2. Overall Rating for Attention to Coordination

We combine our ratings for meeting targets and for relative performance in aid coordination and alignment with qualitative information from donor reports and peer reviews to assign an overall rating of donors' level of attention to coordination, displayed in column three of *Figure 6*. While all donors have a stated aim of increasing "aid effectiveness" in terms of the indicators laid out by the Paris Declaration, donors with higher ratings place a greater emphasis on coordination as a measure of their aid results and performance.

Three donors who rate highly compared to targets and other donors, and for whom there is additional evidence of strong attention to coordination, are rated "high" for attention to aid coordination results. Danida meets the 2011 OECD targets on all indicators of coordination and alignment between donor and recipient aid systems and is ranked fourth

---

[18] For bilateral donors, the data represent results across all country organizations providing ODA, and not just the bilateral organizations we are focusing on for this review. As a result, overall country performance may not accurately represent attention to coordination and alignment by particular bilateral organizations, even though the organizations we review are responsible for the majority of their countries' ODA.

[19] The Global Fund and GAVI do not have data for certain indicators within this report.

[20] This survey covers a varying proportion of each donor's total country programmed aid, with 70 percent as the mean proportion of total official development assistance represented in the survey. Thus, the results reported in the review may not be representative of a donor country's global aid efforts. In particular, the countries included in the survey comprised less than half of the total ODA for AFD, ADC, and the IMF.

[21] Targets for coordinated technical support, joint missions, and joint analyses were the same for all donors, while other targets were different for each donor and were calculated based on 2005 baseline measures. These targets are set by the OECD and may not reflect individual targets established by donors or partner countries.

[22] The sub-index for avoidance of parallel implementation structures is an exception, as in this case higher values for the number of PIUs represent worse performance. Thus, this sub-index is calculated using the formula *index = (actual - maximum) / (minimum - maximum)*.

on the aid effectiveness index. Their 2011 results framework (Ministry of Foreign Affairs of Denmark, 2011) stresses that partner systems should be used to generate results. Sida met the 2011 targets for most indicators and improved its performance since 2008 on nearly all indicators. The Dutch MFA do not meet the OECD 2011 targets for some indicators, but approach the targets in all cases and show improvements since 2008. The Dutch MFA is ranked third among donor organizations in aid effectiveness based on 2008 data, indicating that its targets for performance in coordination in 2011 may have been higher than those of other organizations.

Five donors (ADB, GIZ, IMF, JICA, WBG) are rated "medium-high." For example, Germany has reorganized its development organizations into a single entity, GIZ, and has significantly improved its performance in cooperation (BMZ, 2013). The IMF is rated "medium-low" in meeting aid effectiveness targets but is not far below those targets, and a 2014 review of IMF technical assistance finds "improved country ownership and coordination with donors" (Independent Evaluation Office of the IMF, 2014e).

Four donors (AFD, DFID, DFAT, GAVI, IADB) are rated "medium". A 2013 OECD peer review finds that AFD (France) "to a very broad extent makes use of national procedures" and "is involved actively in the policy dialogue and in aid coordination mechanisms." A separate 2013 OECD peer review finds that "Australia [DFAT] has made good progress with delivering aid as defined in the Busan Partnership for Effective Development" although there is still room for improvement. For the GAVI Alliance, a 2009 report (HLSP, 2009) finds that GAVI systems were generally parallel to national systems, but a 2011 report (Vivideconomics, 2011) finds improvements in developing recipient-country ownership.

Six donors (ADC, AfDB, DFATD, Global Fund, Norad, UN Group) are rated "medium-low," and in the cases of DFATD, Norad, and the UNG, this rating is influenced by other evidence besides the OECD (2011) and Sinha (2010) reviews. A 2012 OECD peer review of DFATD (Canada) finds that it is "lagging behind in implementing the aid effectiveness principles," due in part to more concern with internal efficiency than the principles themselves. The performance of Norad on the aid effectiveness indicators worsened between 2008 and 2011. A 2014 report finds that Norad sought to limit the "administrative burden imposed on partners," but that this often led to reduced coordination as Norad did not provide much support or guidance to its partners (ITAD, 2014). The UNG does not meet its targets for most indicators of aid effectiveness, but it approaches them, and DFID's 2011 Multilateral Aid Review found that "Almost all of the UN organizations are satisfactory or strong on partnership behavior."

Just one donor, USAID, is rated "low" for attention to coordination. Very little of USAID aid used recipient-country systems, and USAID also has a large number of parallel PIUs, with 448 across 61 Paris Declaration survey countries in 2010. While it appears to be making some efforts to coordinate technical assistance, donor missions, and recipient-country analyses, it still performs perform poorly in these areas. USAID is also ranked 29th on Sinha's aid effectiveness index, the lowest of the countries we review.

4.3.3. Coordination Challenges

External reviewers of donor organizations tend to focus on areas where donors face challenges or have opportunities to improve, and many of the external reviews pay particular attention to donor and recipient alignment and coordination. Value judgements made by these external reviewers may bias the selection of challenges presented here. In the review spreadsheet sent in conjuction with this report, we therefore list the external documents we reviewed in a separate column from the other evidence we considered to highlight a potential reason why our review would have uncovered more challenges than successes for particular donor organizations.

Of the 22 donor organizations we review, 13 report challenges for coordination of results measurement with grantees or implementing partners, as displayed in column four of *Figure 6*.These challenges generally fall into three categories: taking action to improve coordination, coordinating with numerous diverse stakeholders, and coordinating joint evaluations.

Four organizations (Norad, IADB, Sida, WBG) report a need for **additional actions to improve coordination**. For example, in the case of Norad (Norway), "planning and execution is left to the partners" and "the majority of grants do not have good results frameworks and cannot be evaluated for outcomes" (ITAD, 2014). A review of the IADB finds that "some partners [are] concerned about bureaucracy and limited decentralization" (DFID, 2011).

Five organizations (AFD, ADB, Global Fund, JICA, ADC) describe difficulties with **coordinating results measurement with diverse stakeholders** involved. A 2013 OECD peer review of AFD finds that "The number and diversity of all stakeholders, some of whom partially cover the same sectors, does not facilitate the co-ordination and the consistency of French action"

(OECD, 2013a). A 2009 evaluation of the Global Fund reports challenges with developing common protocols, indicators, and measurement tools for data collection and analysis as a result of working with numerous partners in the countries in which it operates (Macro International Inc, 2005). A 2014 external review of JICA opines that Japan's aid policies could more representatively reflect the diversity of recipient countries (International Development Center of Japan, 2014).

In addition, four organizations (GAVI, USAID, DFID, GIZ) report challenges in **coordinating joint evaluations** with their partners. A 2013 evaluation of USAID (United States) evaluations finds that 65 percent of evaluation teams lacked members from local populations (Hageboeck, Frumkin, & Monschein, 2013). A 2014 OECD peer review finds that "DFID has not managed to reduce its reporting requirements, and programming processes have become even more cumbersome" (OECD, 2014). A 2011 review of several multilateral organizations reports that in GAVI, "beneficiaries do not have a direct voice in the governance structure" (Vivideconomics, 2011).

Four organizations provide examples of **actions to improve coordination** or to address challenges in alignment of results measurement with partners. The GAVI Alliance's Full Country Evaluation emphasizes the principles of "work[ing] collaboratively with country partners to conduct primary data collection," "partnering with in-country institutes and undertak[ing] shared dissemination," and "Provid[ing] timely, regular, and systematic feedback to country institutes" (IHME, 2013). The Global Fund has established Country Coordinating Mechanisms (CCMs) that are considered to be broadly successful in improving aid coordination (Vivideconomics, 2011). A 2014 OECD peer review finds that DFID "has engaged in a comprehensive review and reform process aimed at streamlining its procedures" and is "a strong supporter of joint programmes and approaches," though its focus on value for money and results has limited its ability to coordinate with partners (OECD, 2014). Finally, DFATD has implemented a new Authorized Programming Process (APP), "which standardizes programming procedures to increase efficiency" and ease coordination burdens (DFATD, 2014a).

4.4. Results - Outputs and Implementation

Donors often assess performance, from the project level to the donor level, using output results. "Outputs" are the direct results of a project or program. Examples of indicators for measuring outputs include numbers of insecticide-treated bed nets delivered, numbers of teachers trained, and kilometers of roads or railways constructed.

In this section, we review output results frameworks for bilateral and multilateral donors. *Figure 7* is a heat map summarizing our findings. It displays donors' level of attention to specified output indicators and their methods for assessing output results. We use annual reports and external reviews as strong evidence for the collection of output data. When internal planning or guidelines documents are the only available source of information, questions are coded as "planned."

> **Key Findings on Measuring Output and Implementation Results**
> - Over three quarters of both bilateral and multilateral donors assess performance in terms of outputs of activities.
> - Multilateral donors are more likely than bilateral donors to specify output indicators and actually collect data based on those output indicators.
> - Multilateral donors are more likely than bilateral donors to collect and report gender-disaggregated output data.
> - Poor output data quality is a challenge for some donor organizations, affecting assessment throughout the results chain. External reviews indicate that better development of results frameworks could support improved output data collection.

We find that nine of 12 bilateral donors and eight of ten multilateral donors assess performance in terms of outputs from their projects (column one in *Figure 7*). Two multilateral donors (GAVI Alliance and UNG) and two bilateral donors (Norad and USAID) "somewhat" assess performance by outputs. In all four of these cases, external reviewers find that reporting on outputs occurs for some projects, but is not comprehensive. For one donor (MFA), our documents do not provide evidence on assessing performance in terms of outputs.

Figure 7: Measuring Output and Implementation Results
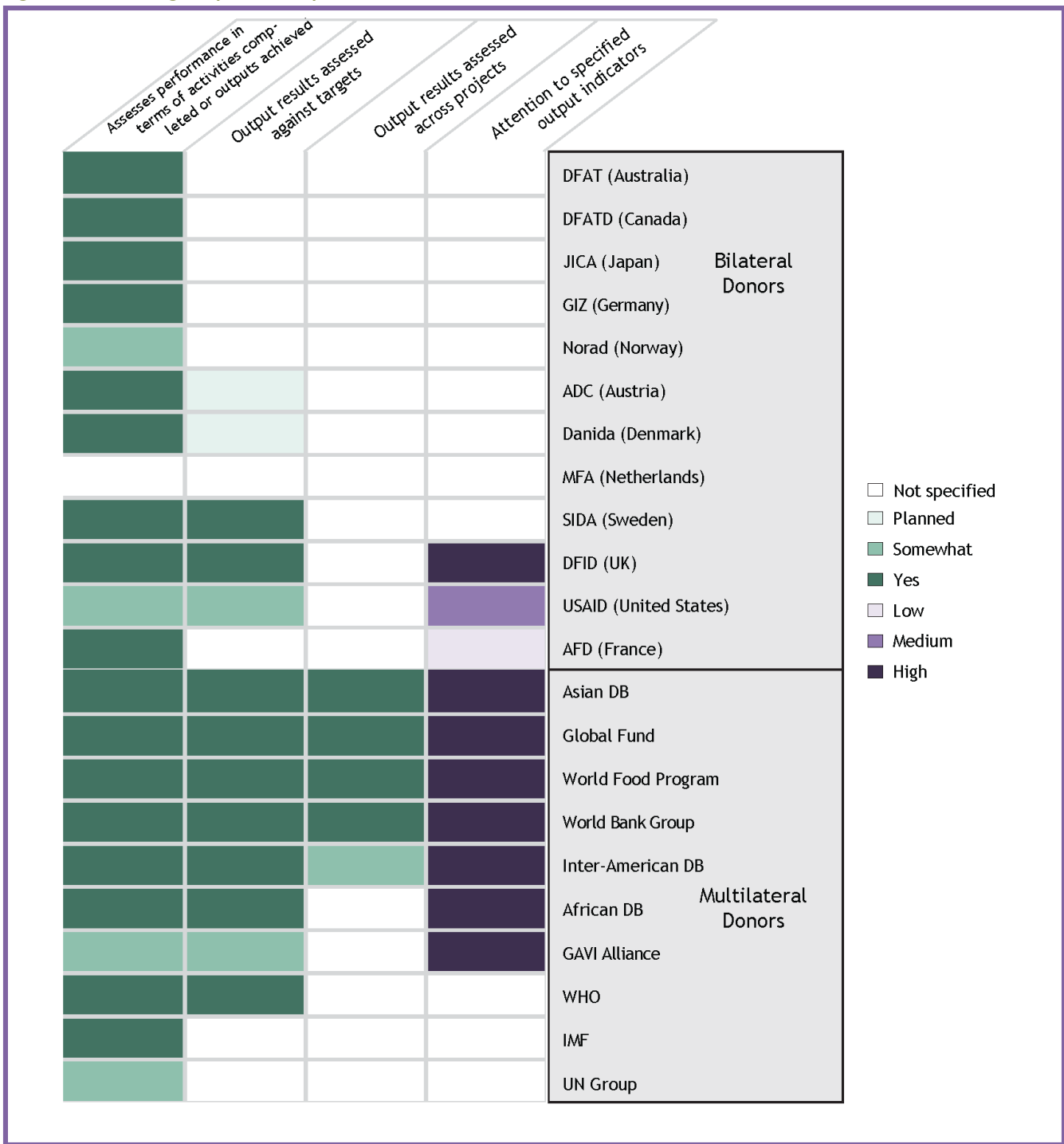
4.4.1. Specification of Output Indicators

We find that four of 12 bilateral donors (AFD, USAID, DFID, Sida) and seven of ten multilateral donors specify an organization-wide list of output indicators in order to use the same indicators across projects. Donors that assess performance in terms of outputs but do not have a list of output indicators may choose different output indicators for each project. For example, an internal audit of DFATD (Canada) notes that there is not an organization-wide list of output indicators. Instead, data are collected on around 8,000 indicators, many of which are similar (Canadian Chief Audit

Executive, 2012b). When data are collected on slightly different output indicators, data aggregation and comparison is difficult.

We rate attention to specified output indicators as "low," "medium," "high," and "not specified," as displayed in column four of *Figure 7*. For a "low" rating, there is some evidence that the donor specified a list of output indicators, but little or no evidence that these output indicators are used consistently across projects. For a "medium" rating, output indicators are specified and there is evidence that data based on these indicators are actually collected, but there are noted problems with data collection, quality, or comprehensiveness. For a "high" rating, the donor specifies a list of output indicators, and there is evidence that data on these indicators is actually collected and aggregated (e.g. aggregated figures on output indicators in an annual report, or an external reviewer reporting on the donor's results management system). For "not specified" donors, there is no evidence in the documents we review that the organization specified consistent output indicators across projects.

Of the 12 bilateral donors, we rate AFD "low," USAID "medium," and DFID "high." The remaining nine bilateral donors are "not specified." DFID is the only bilateral donor that specifies a list of output indicators and demonstrates that data from these indicators were actually collected consistently. In its 2014 annual report, DFID reports aggregated figures for output indicators, including numbers of people benefiting from cash transfer programs supported by DFID and number of teachers trained through DFID programs (DFID, 2014a). USAID also reports aggregated output data, but an external reviewer has criticized the comprehensiveness of its results management system. Lawson (2013) finds that there are limited resources allocated to results measurement and evaluation, and that results measurement is focused more on "tracking where funds go and what they have purchased" rather than the outputs, outcomes, or impact of funds. He adds that "ad hoc evaluations of limited scope and learning value" are often the default approach due to staffing and budgeting limitations, an emphasis on controlling evaluation cost, and a lack of requirement for well-designed evaluation plans. For AFD, there is an official list of indicators, but no evidence in the documents we reviewed that this list is used across projects.

In contrast, we rate attention to specified output indicators as "high" for seven of ten multilateral donors, and "not specified" for three (WHO, IMF, UNG). All three regional development banks (ADB, IADB, AfDB) have specified lists of indicators, and report aggregated data for these indicators in annual reports. Similarly, the WFP aggregates and reports numbers of beneficiaries and amount of food distributed across projects (WFP, 2014). For the WHO, IMF, and UNG, we did not find an organization-wide list of output indicators.[23]

Some donors also specify that output data for appropriate indicators should be gender-disaggregated. For example, instead of reporting the number of farmers participating in an extension program, donors may require that data on men and women farmers are collected and reported separately. For **attention to gender-disaggregated output indicators**, we use the same rating criteria as for attention to specified output indicators. Of the bilateral donors, only DFID, which we rated "medium," has any evidence of collecting and reporting gender-disaggregated output data. In its annual report, DFID reports gender-disaggregated output data for some indicators, including people with access to financial services through DFID's activities. Other output indicators which could have been gender-disaggregated, including number of teachers trained, were not. No other bilateral donors consistently report gender-disaggregated output data in any documents we reviewed, or state that reporting gender-disaggregated data was an official policy.

For the ten multilateral donors, we rate three as "high" (ADB, IADB, AfDB), four as "low" (WFP, Global Fund, GAVI, UNG), and the remaining three as "not specified." For all three of the regional development banks, annual reports include gender-disaggregated output data for appropriate indicators. For the four multilateral donors we rated as "low," the organization specifies that gender-disaggregated data should be collected, but there is no evidence that gender-disaggregated data are actually collected and aggregated. GAVI, for example, specifies that gender-disaggregated data should be collected in its gender policy document, but reported figures on vaccine recipients are not gender-disaggregated (GAVI, 2012; GAVI, 2013b).

4.4.2. Assessing Output Results

Donors have several methods for using output data to assess performance. Donors may set targets for output indicators at the project level and/or organizational level, and **assess performance in terms of meeting output targets**. We find

---

[23] For donors that specified output indicators, output indicator categories and specific indicators varied. For health-focused multilaterals (GAVI Alliance and Global Fund), output indicators included numbers of people receiving various vaccines (GAVI Alliance) and number of malaria nets distributed (Global Fund). Development banks (WBG, ADB, AfDB, IADB) used similar output indicators, including kilometers of roads constructed, and numbers of people participating in their programs.

evidence that two bilateral donors (Sida, DFID) consistently assess performance in terms of achieving output targets. Although USAID sets targets for output indicators, there are criticisms of the comprehensiveness of its overall results management system due to its focus on accountability of funds over outputs, outcomes, or impacts (Lawson, 2013), so we rate it as "somewhat" for assessing performance against output targets. For two bilateral donors (Sida, ADC), planning documents state that output results should be assessed in terms of meeting targets, but we find no evidence that these targets are actually used.

Assessing **performance in terms of meeting output targets** is more common among multilateral donors, as displayed in column two of *Figure 7*. Seven of ten multilateral donors assess performance in terms of outputs achieved against targets. For example, at the project level, the WFP sets output targets for number of people reached through a project, and then uses those targets to assess its performance. The ADB sets output targets at the aggregated project-level (number of beneficiaries reached) and at the organizational level (number of projects completed and rated "satisfactory"). We rate the GAVI Alliance as "somewhat" assessing output results in terms of targets, due to the criticism of the comprehensiveness of its results management system (GAVI, 2011). For two multilateral systems (UNG and IMF) we find no evidence that they assess performance in terms of achieving output targets or plan to begin using output targets.

Another method for assessing output results is **comparing output results across projects**, displayed in column three of *Figure 7*. If comparable indicators and results measurement systems are used, donors can compare projects in terms of outputs achieved. We find that none of the 12 bilateral and four of ten multilateral donors (ADB, WFP, World Bank Group, Global Fund) have processes for comparing output results across projects. For example, the WFP compares the number of beneficiaries reached in different projects (WFP, 2013), and an external review of the WBG found that the organization's Independent Evaluation Group uses standardized indicators and reporting forms to compare results across projects (Lloyd et al., 2014). We rate the IADB as "somewhat," because there are specific examples of comparisons of output results across projects in its annual report, but there is no evidence that comparisons are made systematically.

4.4.3. Challenges with Output Level Results

Challenges for collecting information about outputs are identified for two of 12 bilateral donors (Norad, USAID), and for three of 10 multilateral donors (ADB, WFP, UNG). Poor data quality is noted as a challenge for three donors (WFP, UNG, Norad). For Norad, an inadequate framework for collecting data is noted, since many project-level evaluations did not identify outputs at all (Norad, 2014). For the WFP, an internal review noted that the organization's output indicators, including number of beneficiaries reached, are not specific enough to provide high quality data for assessing projects (WFP, 2014). Although there are differences across UNG organizations, DFID's 2011 Multilateral Aid Review found that data quality is often poor due to an inadequate results framework.

4.5. Results - Outcomes and Impact

In addition to measuring outputs, donors also often assess their performance in terms of outcome-level results. "Outcomes" are desired results caused by outputs. For example, delivery of vaccines, an output, may cause changes in the recipient-country's child mortality rate, a desired outcome. The Millennium Development Goals are often used as a framework for choosing outcome-level indicators and targets, including rates of child mortality, rates of stunting and wasting, and indicators of gender equality. Unlike output results, to assess performance by an outcome measure donors must make some effort to attribute those outcomes to programs activities. For example, while the number of vaccines delivered or administered is a direct output or activity measure of performance, changes in the child mortality rate could be partly or wholly attributable to other simultaneous changes, such as improvements in water quality, nutrition, or neo-natal and child health care systems.

> **Key Findings on Measuring Outcome and Impact Results**
> - Just over half of bilateral and multilateral donors assess performance in terms of outcomes achieved, fewer than consider outputs.
> - Multilateral donors are more likely to specify outcome indicators and show evidence for collecting data based on those indicators.
> - Common challenges include lack of baseline data and other data quality issues, and an organizational structure that focuses on outputs or processes over outcomes.

In this section, we review outcomes results frameworks for bilateral and multilateral donors. Column one in *Figure 8* is a heat map summarizing our findings, where bilateral and multilateral donors with similar qualities are clustered together. Columns two, three, and four display different methods for how donors assess performance in terms of outcomes. We find

that **six of 12 bilateral donors and six of ten multilateral donors assess performance in terms of outcomes achieved from their projects**. Two bilateral donors (JICA, Danida) "somewhat" assess performance in terms of outcomes.

Figure 8: Measuring Outcome and Impact Results

	Assesses performance in terms of outcomes	Outcome results assessed against targets	Outcome results assessed against control groups	Attention to specified outcome indicators
DFID (UK)	Yes	Yes	Not specified	High
DFAT (Australia)	Yes	Yes	Not specified	High
USAID (United States)	Yes	Yes	Yes	High
GIZ (Germany)	Yes	Not specified	Yes	High
AFD (France)	Yes	Not specified	Yes	Low
SIDA (Sweden)	Planned	Planned	Planned	High
Danida (Denmark)	Somewhat	Planned	Planned	Medium
ADC (Austria)	Not specified	Not specified	Not specified	Low
MFA (Netherlands)	Not specified	Not specified	Not specified	Low
JICA (Japan)	Somewhat	Not specified	Not specified	Low
DFATD (Canada)	Yes	Yes	Not specified	Not specified
Norad (Norway)	Not specified	Not specified	Planned	Not specified
Inter-American DB	Yes	Yes	Yes	High
Asian DB	Yes	Yes	Yes	High
World Food Program	Yes	Yes	Yes	High
World Bank Group	Yes	Not specified	Yes	High
GAVI Alliance	Yes	Yes	Not specified	High
Global Fund	Yes	Yes	Not specified	High
African DB	Somewhat	Somewhat	Not specified	High
UN Group	Planned	Not specified	Planned	Low
WHO	Planned	Not specified	Not specified	Not specified
IMF	Somewhat	Not specified	Not specified	Not specified

Bilateral Donors (DFID through Norad); Multilateral Donors (Inter-American DB through IMF)

Legend: Not specified / Planned / Somewhat / Yes / Low / Medium / High

4.5.1. Specification of Outcome Indicators

Ten of 12 bilateral donors and eight of ten multilateral donors specify organization-wide outcome indicators. In many cases (DFAT, Danida, JICA, MFA, DFID, UNG), donors use outcome indicators based on the Millennium Development Goals (MDGs). Other common categories of specified outcome indicators included gender equality (DFAT, AfDB, ADB, ADC, GIZ, WBG) and environmental sustainability (AfDB, ADB, IADB, ADC, GIZ).

Similar to our ratings for output indicators, we rate attention to specified outcome indicators as "low," "medium," "high," and "not specified," as displayed in column four of *Figure 8*. A rating of "not specified" indicates that we found no evidence that the donor specifies a consistent list of outcome indicators. A "low" rating indicates that the donor specifies a list of outcome indicators, but we find no evidence that the donor actually consistently collects outcome data based on the specified indicators. For "medium" donors, there are specified outcome indicators and the donor collects outcome data based on the indicators, but there is evidence of issues with data collection, quality, or comprehensiveness. A "high" rating indicates that there is evidence that the donor specifies outcome indicators and systematically collects outcome data based on these indicators.

Out of 12 bilateral donors, five receive a "high" rating (Sida, GIZ, DFID, USAID, DFAT), one receives a "medium" rating (Danida), four receive a "low" rating (JICA, AFD, MFA, ADC), and two are "not specified" (Norad, DFATD). DFID is an example of a "high" rated donor, which states that it tracks outcomes related to the MDGs, and reports figures for these outcome indicators in its annual report (DFID, 2014a). Danida is rated "medium" because while it tracks outcome indicators related to the MDGs, an external review finds that outcome data for Danida are inconsistent. JICA is an example of a "low" rated donor, since it states that it tracks outcomes related to the MDGs, but an external reviewer finds that outcome indicators are usually absent from JICA's evaluations, and are not aggregated in JICA's annual evaluation reports (International Development Center of Japan, 2014a).

For the 10 multilateral donors, seven are rated "high," (WBG, WFP, IADB, ADB, Global Fund, GAVI, AfDB), one is rated "low" (UNG), and two are "not specified" (IMF, WHO). As an example of a highly rated donor, the WFP specifies outcome indicators and reports figures for stunting rates and school enrollment in countries where it works. The UNG is rated "low" because while it specifies that MDGs are a priority across the UN system, there is no evidence that MDG-based outcome data are collected system-wide.

4.5.2. Assessing Outcome Results

Donors assess outcome-level data using several methods. One method is **setting targets for outcome-level results**, and then assessing performance against those targets. We find that four of 12 bilateral donors (DFID, DFATD, DFAT, USAID) and six of ten multilateral donors (IADB, ADB, WFP, GAVI, Global Fund) consistently assess performance in terms of meeting outcome-level targets. For example, the IADB's evaluations assess changes in outcomes against targets, and assess how attributable those changes in outcomes are to the IADB's programs (IADB, 2014b). Other donors, like DFID, have donor-level outcome targets used to assess overall organizational performance (DFID, 2014a). The AfDB is rated as "somewhat" assessing performance in terms of outcome targets, since some, but not all, of the evaluations we reviewed include outcome targets. For two bilateral donors (Danida, Sida), information in planning documents states that the organization tracks outcomes against targets, but it is not clear if this is actually carried out.

Another method for assessing performance using outcome-level results is **comparisons of performance against baseline data**. We find that two of 12 bilateral donors (USAID, DFATD) and four of 10 multilateral donors (WBG, AfDB, GAVI, IADB) systematically use baseline data to assess performance in terms of changes in outcome-level results. As an example, DFATD uses changes from baseline measures of school attendance to assess its performance. Two of 12 bilateral donors (AFD, DFAT) "somewhat" assess performance in terms of changes to outcome-level results. In both cases, there are specific examples of the use of baseline measures, but no evidence that baseline measures are used systematically. Two multilateral donors (WFP, ADB) also "somewhat" used baseline measures. These two multilaterals were rated "somewhat" due to evidence of data quality issues. Finally, three donors (WHO, Norad, Sida) discuss using baseline measures in planning documents only, while the remaining donors are rated "not specified." These donors may collect and use baseline data, but we find no descriptions of baseline data in our body of evidence.

Another method for assessing performance using outcome-level results is to **compare changes in outcomes across projects**. In this case, some method for attributing changes in outcomes to a project is needed. Only two donors (WBG, ADB) display evidence of regularly making this type of comparison. For example, in the ADB's case projects are rated "satisfactory" or "unsatisfactory" based on contributions to ADB's outcome-level priorities, including environmental sustainability and gender equality (ADB, 2015).

Finally, donors may also **use control groups to assess outcome-level results**. Comparing changes in control groups (not receiving aid or the intervention) relative to treatment groups (receiving aid or the intervention but otherwise resemble the control group) provides additional confidence in attribute treatment group changes to the donors' programs. Five of

ten multilateral donors (ADB, AFD, USAID, WBG, WFP) specify using control groups, though only two report frequent use of control groups, WBG and ADB. ADB has 24 ongoing impact evaluations (as of 2013), and the WBG has 81 percent of ongoing evaluations using experimental methods. For the remaining three multilateral donors, it is not clear how large a share of their evaluations use control groups. Two bilateral donors (USAID, AFD) use control groups, but to a limited extent. For AFD, we find no evidence for how often control groups were used, and for USAID only 3 percent of ongoing evaluations were using control groups between 2008 and 2012 (USAID, 2013d). Four other donors (Norad, Danida, Sida, UNG) mention using control groups in planning documents, but it is not clear if these plans are being implemented.

4.5.3. Challenges with Outcome-Level Results

Challenges for collecting outcome-level results information are identified for ten of 12 bilateral donors and for six of ten multilateral donors. Issues with the quality of baseline data are the most commonly identified challenge, affecting four donors (AFD, GIZ, SIDA, ADB). General data quality is identified as a challenge for two other donors (Danida, WFP). Another common challenge for collecting outcome-level data is an organizational focus on outputs over outcomes, identified for two donors (JICA, DFID), or a focus on processes over outcomes (ADC). Two donors (DFAT, UNG) are criticized in peer reviews for low levels of analysis of long-term outcomes. As a more extreme case, Norad's evaluations are criticized for not mentioning outcome-level results at all (Itad, 2014a).

Even when results frameworks are adequate, inherent difficulties in collecting outcome-level results information are also mentioned. For DFATD, the inherent difficulty in attributing changes in outcomes to the donor's programs was mentioned as a challenge, and the ADB noted that it is more difficult to measure outcome-level results for some sectors (e.g. tourism).

<u>4.6. Results - Costs and Effectiveness</u>

Donors measure and report costs with varying levels of attention and analysis. In this section, we report on four categories of performance assessment related to cost: simply reporting costs/spending, comparing costs against budgets, comparing costs against outputs (efficiency), and comparing costs against outcomes (cost-effectiveness). As shown in *Figure 9*, each level of cost results measurement builds on the one before, increasing in complexity, but also in performance information. While reporting costs can increase aid transparency, assessing aid efficiency and cost-effectiveness generally requires a "per unit" analysis of costs: for example comparing costs to outputs or outcomes, or between projects and programs.

> **Key Findings on Measuring Cost and Effectiveness Results**
> - Donor organizations appear to focus much more on "how much we spent" as compared to what was achieved with the money spent.
> - Few donors compare costs against outcomes to assess cost-effectiveness, citing inherent difficulties in accurately measuring outcomes.

Figure 9: Hierarchy of Cost Measurement
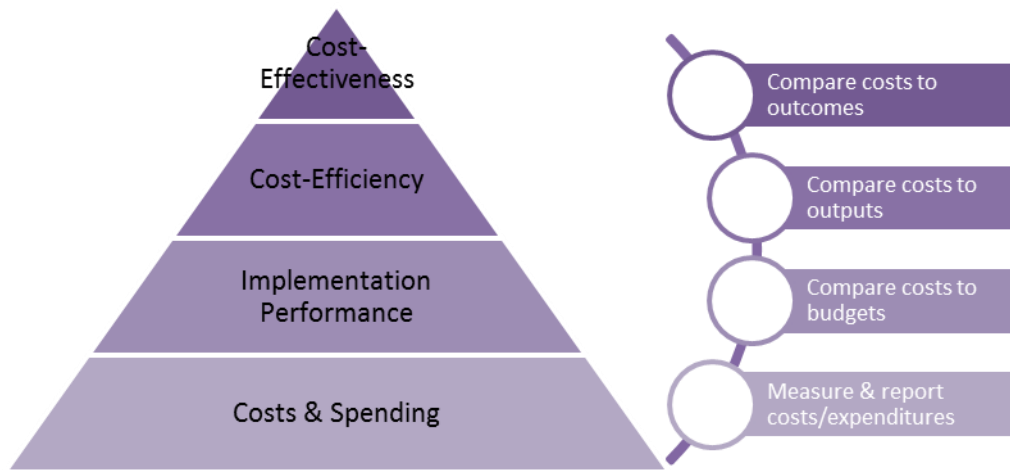
The heat map in *Figure 10* presents the attention given to cost measurement. For the four categories, we rate donors' attention to using cost information to assess performance. Donors who receive a "high" rating appear to pay attention to assessing performance with cost information, with evidence provided by external reviews or internal evaluation reports.

Donors receive a "medium" rating when they report or assess cost information for some but not all projects. For example, internal guidelines or frameworks may state that the donor uses costs to measure performance, but reports may not always include this information. A "low" rating means the donor appears to give little attention to cost information; internal documents may suggest limited attention to cost evaluations, or external evaluations suggest room for improvement.

Figure 10: Measuring Cost and Effectiveness Results

	Assess performance in terms of costs	Assess performance in terms of costs against budgets	Assess costs against outputs	Assess costs against outcomes
Bilateral Donors				
DFID (UK)	High	Medium	High	Low
USAID (United States)	Medium	Medium	Medium	Low
GIZ (Germany)	High	Not specified	Medium	Medium
DFAT (Australia)	High	Medium	Not specified	Low
DFATD (Canada)	High	High	Not specified	Low
SIDA (Sweden)	High	Medium	Not specified	Not specified
Danida (Denmark)	High	Not specified	Not specified	Not specified
AFD (France)	High	Not specified	Not specified	Not specified
Norad (Norway)	High	Not specified	Low	Not specified
JICA (Japan)	Low	Not specified	Not specified	Low
ADC (Austria)	Not specified	Not specified	Not specified	Not specified
MFA (Netherlands)	Not specified	Not specified	Not specified	Not specified
Multilateral Donors				
World Bank Group	High	Not specified	Not specified	High
GAVI Alliance	High	Not specified	Not specified	High
Inter-American DB	High	High	Medium	High
Global Fund	High	High	Not specified	Medium
African DB	High	High	Not specified	Not specified
IMF	High	Not specified	High	Not specified
UN Group	High	Not specified	Medium	Not specified
World Food Program	High	Not specified	Not specified	Not specified
Asian DB	Medium	Not specified	Not specified	Not specified
WHO	High	Not specified	Low	Not specified

Legend: Not specified; Low; Medium; High

4.6.1. Measuring Spending on Interventions

As shown in column one in *Figure 10*, **ten donors receive a "high" and nine receive a "medium" rating for assessing performance in terms of the amount of spending on donor-funded projects**. Donors with "high" ratings explicitly report

using costs to assess performance, while those with "medium" ratings report costs but do not specify how costs are used to measure performance. In general, donors measure and report costs in terms of allocations to different projects, countries, or sectors, or as reports on total expenditures. Project costs and total spending are often reported in annual reports, with expenditures for particular programs or activities often presented as results of donor activities. For example, annual reports for DFAT include the amount of funding allocated to different programs, the AfDB's annual reports display amounts approved and disbursed as grants and loans, and AFD's reports include spending and allocations by project and recipient-country. In all three examples, spending on projects is presented as a primary measure of organizational performance. Some organizations, however, are criticized for over-emphasizing reporting on spending as opposed to outputs, outcomes, or impact. For example, Lawson (2013) finds that at USAID focuses on "tracking where funds go and what they have purchased rather than the impact of funds on development or strategic objectives."

4.6.2. Comparing Costs Against Budgets

Fewer donors (8 of 22) receive "high" or "medium" ratings for comparing costs or expenditures against budgets to assess implementation performance, as shown in column two in *Figure 10*. While some donors simply listed budgeted amounts next to expenditures, others assessed the proportion of project spending that was in line with budgets. This type of results measurement is primarily used to evaluate organizations' implementation performance, and to support accountability of the organizations to their stakeholders.

The Canadian DFATD is exemplary; their 2014 Department Performance report breaks down planned and actual spending by program and sub-program and annotates reasons for discrepancies between planned and actual spending (DFATD, 2014a). The ADB reports the percentage of projects with cost overruns in their annual Development Effectiveness Review. They report that in 2013, for example, about 16 percent of projects with successful completion reports exceeded their budgets or had to reduce outputs to stay in budget (ADB, 2013a). Budget monitoring may be internal or external. For example, the Office of Inspector General conducts audits of USAID program operations, including financial statements and performance in implementing programs (USAID, 2013e). Other donors appear to monitor budgets without explicitly mentioning using budget/spending discrepancies to assess implementation performance. For example, the National Audit Office conducted audits of DFID's financial statements (OECD, 2014a), but the documents reviewed do not specify how results were used.

4.6.3. Comparing Costs Against Outputs Achieved

In addition to comparing costs to budgets, donors may assess implementation performance and **cost-efficiency** by comparing costs to the outputs achieved. Programs which produce more outputs compared to amount spent are considered more cost-efficient. Three agencies (Global Fund, DFID, and GIZ) are rated "high" for measuring costs against outputs (column three, *Figure 10*). The Global Fund, through its use of performance-based financing, requires grantees to report costs and results in order to receive continued funding. Indicators in their M&E framework include the price of outputs (Global Fund, n.d.). DFID's Annual Report (2014) includes indicators such as "cost per child supported in primary education" and "average unit price of long-lasting insecticide treated bed-nets procured." In addition, the United Kingdom's National Audit Office conducts value for money reviews of DFID programs (OECD, 2014a). GIZ (Germany) rates project/program efficiency on a six-point scale, defining efficiency as "Are the objectives being achieved cost-effectively? In other words, are the resources invested in a development measure (funding, expertise, time, etc.) appropriate compared to the outputs and results achieved?" (GIZ, 2013). However, the same document acknowledges the challenge of measuring efficiency.

Three donors (AfDB, USAID, and IADB) are rated "medium" and four (Norad, DFAT, JICA, and WFP) are rated "low" ratings for comparing costs to outputs. The WFP, which receives a "low" rating, describes plans to improve cost reporting by aligning its performance plan with the budgeting process. Their 2014 plan requires divisions or units to present a results statement and list the costs of all contributing activities (MOPAN, 2013). In some cases, donors indicate commitments to achieving results efficiently or providing "value for money", but do not outline how these commitments would be evaluated. For example, DFAT's framework describes a standard for ensuring value for money, but does not outline how it would be calculated. An internal review found that few evaluations compared costs to outputs or conducted full "value for money" analyses (DFAT Office of Development Effectiveness, 2014). The AfDB, which receives a "medium" rating, has committed to keeping "costs under constant scrutiny and make sure that all our resources—human and financial—are used efficiently and economically to deliver maximum development impact" (AfDB, 2013c). They appear to compare costs against outputs at the organizational level but not on the individual project or sectoral level. Their Development Effectiveness Review (AfDB, 2014b) measured cost efficiency through several indicators, including administrative costs per

UA[24] 1 million disbursed, cost of preparing a lending project, and cost of supporting project implementation, but there is no mention of costs per output provided to beneficiaries and no other reports appear to measure cost efficiency.

4.6.4. Comparing Costs Against Outcomes Achieved

To assess the **cost-effectiveness** of programs, certain donors compare costs to realized outcomes. Programs that achieve desired outcomes at a lower cost are considered more "cost-effective", but as indicated in Section 4.5.3, difficulties in measuring outcome indicators make measuring cost-effectiveness of programs at the outcome level inherently challenging. As shown in column four in *Figure 10,* only three donors (GAVI, IADB, and WBG) are rated "high." For example, GAVI conducts in-depth analyses in a small number of countries to assess the impact and cost-effectiveness of introducing and scaling up coverage of new vaccines. Measures of cost-effectiveness include cost per life saved (GAVI, 2011).

Two donors (GIZ, Global Fund) receive "medium" ratings, and six donors receive "low" ratings. **Half of all donors (11 of 22) do not specify whether they measure cost-effectiveness at the outcome level.** For some donors rated as "not specified," indicators measuring cost-effectiveness may exist but are not reported on in the documents we review. DFAT, for example, described a standard for ensuring value for money, but does not outline how it would be calculated. A review of their aid evaluations finds that few evaluations conducted value for money analyses or engaged in discussion around relative costs of achieving outcomes (DFAT, 2014a).

4.6.5. Challenges with Cost-Based Assessments

Only two donors (WBG, GIZ) appear to compare cost-effectiveness results across projects. An evaluation by the WBG's Independent Evaluation Group indicates that 21 percent of completed independent evaluations include an efficiency analysis, which is comprised of cost-benefit analysis, economic rate of return, or cost effectiveness analysis across different projects or treatment types (IEG, 2012b). GIZ rates project efficiency on a scale of one to six, implying comparison across projects and programs, but it is not clear whether it measures cost-efficiency of outputs or of outcomes.

Donors report that cost-efficiency and cost-effectiveness are difficult to measure (GIZ, WBG, Sida) and document (Norad). Sida reports that these measures are difficult to define because the language does not distinguish between "efficiency" and "effectiveness" (Sida, 2007). An evaluation of the WBG's impact evaluations indicates that cost-effectiveness was given less attention because projects were often funded by multiple sources, making it difficult to account for all the costs of a project (IEG, 2012b). Evaluating cost-effectiveness at the outcome level first requires measurement of results at this level, which can be challenging due to attribution difficulties. Norad's 2012 Annual Report of aid to Afghanistan stated that "cost-effectiveness could not be evaluated because of weak follow-up and documentation." The ADB's results framework included value for money indicators, but they lacked substantial evidence of project-level costs and value (DFID, 2013).

4.7. Use of Results Information

Aid results information can be used for several purposes, including accountability to external stakeholders, institutional accountability and strategic management, and management and improvement of programs, projects, and policy (OECD, 2013c). Some donor organizations use information from results measurement activities to improve overall organizational operations. Others use results information at the program level only. In this section we review organizations' processes for reporting and sharing results and for using results to inform organizational learning, strategy, planning, and funding allocations.

Key Findings on Use of Results Information
- 14 of 22 donor organizations have rules for reporting donor-level results, but the sharing of results internally and externally is not always explicit.
- Multilateral organizations appear to have more developed systems for sharing results information than bilateral donors.
- Most organizations have established systems for using results information for organizational learning, strategy, planning, and improving spending, but use of these systems varies.
- 16 of 22 donors are rated "medium" or "high" for use of results information for strategy and program implementation, but just six of 22 are rated "medium" or "high" for using results to influence funding allocations.

---

[24] Unit of Account for African Development Bank, this is the equivalent of the IMF's special drawing right SDR, the value of which varies daily but is reported in US Dollars.

4.7.1. Systems for Sharing Results Information

Twenty of the 22 donor organizations specify having rules or a plan for reporting results; 14 of these have rules for reporting results at the aggregate organizational level, as shown in Figure 11. Various donor-level reports form the basis for these rules.

The six bilateral organizations **reporting organization-level results** information use a variety of different reports. DFAT (Australia), for example, reports aid performance through Aid Program Performance Reports (APPRs). Canada's bilateral aid organization, DFATD, aggregates results in Departmental Performance Reports (DPRs). DFID (UK), AFD (France), and JICA (Japan) issue annual reports. GIZ develops Corporate Strategic Evaluations in addition to synthesis reports. Eight multilateral organizations report organization-level results information. This information is reported through annual reports (IMF, AfDB, ADB, WHO, WFP), Corporate Results Frameworks (IADB), and in one case through an organization-wide scorecard (WBG).

Of the organizations we review, bilateral organizations are more likely to report results only at the program or project level. Of the six donors (Norad, Danida, USAID, Global Fund, ADC, and Sida) who do not specify rules for reporting results at the donor level but did specify rules for reporting results at the program or project level, five are bilateral organizations (all but Global Fund).

Figure 11: Donor organizations reporting results information
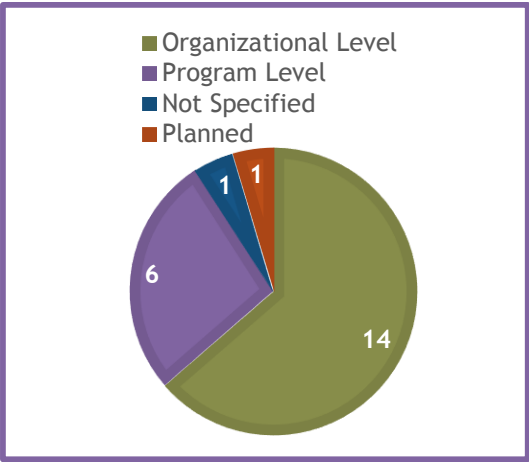
We rate organizations as "high," "medium," or "low" in their attention to sharing information both internally (within the organization) and externally (with grantees, implementing partners, and the public). Bilateral organizations have varying levels of sharing while multilateral organizations showed less variation in attention to internal sharing. As shown in column one of *Figure 12*, we rate three bilateral organizations and one multilateral organization (WBG, GIZ, JICA, AFD) as having a "high" level of **internal results information sharing**, meaning that the organization demonstrates evidence of a culture of sharing results information. For example, the WBG has committed "toward a results culture […] reflected in the corporate scorecard indicators embedded in Managing Directors' performance Memoranda of Understanding" (IEG, 2014b). We further categorized 12 (five bilateral and seven multilateral) organizations as having a "medium" level of attention to internal sharing of results information and two (both bilateral) organizations as having a "low" level of attention. Organizations with "medium" levels of results sharing internally may not explicitly share reports among departments but these reports are posted publicly on a website, which means that the information is available both within the organizations and publicly. Organizations are categorized as "low" if they demonstrate challenges to sharing results information internally. Common challenges include lack of internal communication about lessons learned from successful projects both among institutions and across thematic and recipient-country groups within the donor organizations.

Additionally, 20 organizations describe processes for **sharing results information with the general public**, such as publishing evaluation results on their websites[25]. Two organizations (NORAD, GIZ) specify having presented their results and organizational learning at International Conferences or Seminars.

We also rate donors on their attention to **sharing results information with grantees and implementing partners**. We rate donors as "high" when grantees/implementing partners are involved in evaluation and review processes, as "medium" when regular information sharing occurs (at least posting results information online), and as "low" when there are demonstrated challenges with successful communication with partners/implementers. We classify 15 out of 22 organizations as having "medium" attention to this activity (column two in *Figure 12*). We rate one organization as "high" (AFD) because they have processes for systematically sharing with and receiving feedback from external stakeholders. The three organizations that we classify as having a "low" level of attention to sharing information with implementing partners and grantees each describe challenges to sharing information. For example, a 2014 report found that DFID did not ensure that the implementing farmers "who were the intended beneficiaries" of a program received the research results (Independent Commission for Aid Impact, 2014).

---

[25] Links to websites containing publicly available evaluation reports may be found in the review spreadsheet.

Figure 12: Systems for Sharing Results Information
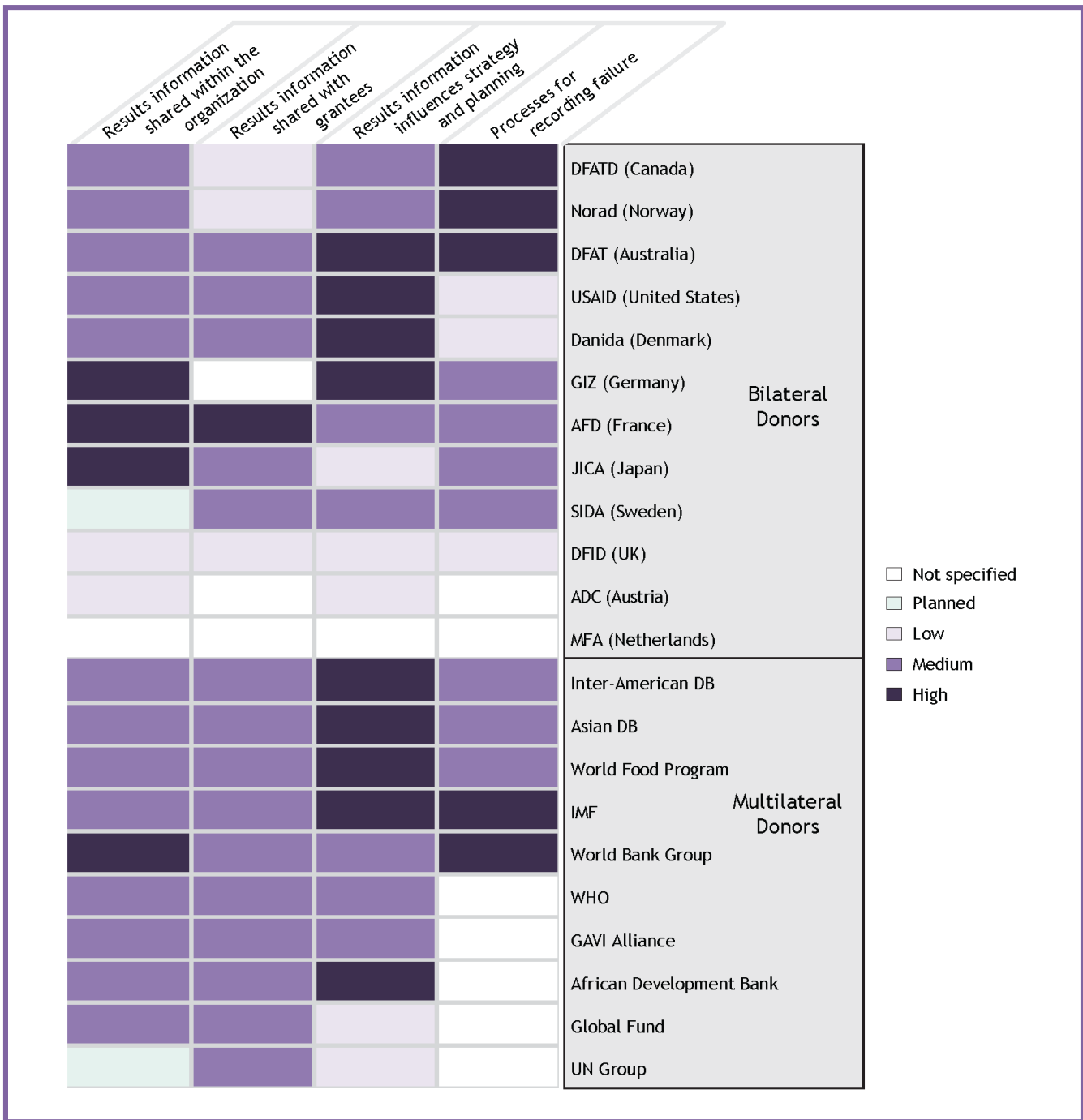
Peer reviews of donor organizations provide some suggestions for improving results sharing. For example, a study of Norad evaluations recommends that Norad "Provide short, clear executive summaries for all reports [and] develop short, non-technical briefs for all reports to make the results more widely accessible" (Grasso, et al., 2012). Other recommendations include more intentional use of seminars, workshops, and conferences for dissemination of results, and quality assurance in transparency. In several cases, organizations receive recommendations to ensure that all evaluations are published (MFA, USAID).

4.7.2. Systems for Incorporating Results Information

Of the 16 donor organizations that describe regularly incorporating results information into their activities, two donors (WBG, DFATD) explicitly indicate the use of an **incentive system to incorporate results**. The World Bank Group has a formal incentives system "for employees to produce positive results" (Lloyd, et al., 2011) and DFATD uses results-based management processes that include compensation measures. The other 14 donors with established processes for incorporating results information describe a variety of processes including **knowledge management documents and policies for organizational learning**. Additional detail on these processes is included in the review spreadsheet that accompanies this report.

Six organizations do not specify whether they have a system for incorporating results information. Of these, four are multilateral organizations. For the two bilateral organizations (Sida, DFID), the documents we review present mixed evidence, indicating that these donors have a process for incorporating results but may not consistently use this process.

Donor organizations vary in the amount of attention they give to the use of results for planning, strategy and project implementation. We rate nine of 22 organizations as "high," meaning that they have systems to ensure that they consistently **incorporate results information into organizational planning and implementation**. For example, a 2014 Peer Review of Danida finds that "A central plank in DANIDA's culture is its 'results orientation'. There is a strong commitment at every level in the organisation that we encountered to emphasise the importance of results and effectiveness" (Kvalvaag & Stern, 2014). A 2013 review of USAID finds that "Three quarters of missions reported that they are using evaluation results to inform project design and improve implementation," though there are difficulties with applying evaluations beyond project implementing teams (USAID, 2013c). A review of the IADB finds that results are used in "the design, appraisal and execution of new operations" (Vivideconomics, 2011).

We rate a further seven organizations as "medium" and five as "low" (column three, *Figure 12*). A "medium" rating indicates that organization documentation mentions the importance of using results information in planning, but there is not a clear process by which this is done, and donors have issues prioritizing information. For example, a 2012 assessment of the use of Norad evaluations found that it "has made a good-faith effort to encourage use of its evaluations" but "that a number of evaluations are not targeted on high-priority issues or are not delivered in time to affect decisions" (Grasso, Imas, & Fostvedt). Organizations rated as "low" demonstrate challenges with using results information or demonstrate that they do not prioritize learning from the results of evaluations. For example, a 2014 review finds that DFID faces challenges "building on experience so as to turn learning into action [...] DFID does not clearly identify how its investment in learning links to its performance and delivering better impact" (Independent Commission for Aid Impact, 2012).
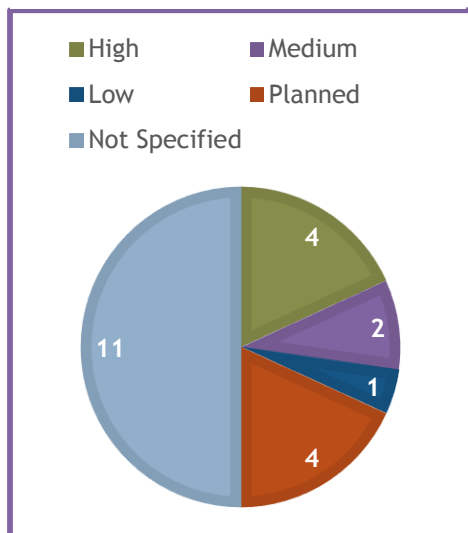
Fewer donors specify **using results to influence budget allocations**. A "high" rating indicates that results information regularly informs budgetary decisions, while donors with a "medium" rating report only using results information sometimes, and organizations with a "low" rating describe numerous challenges in incorporating results information into budget decisions. In four organizations results information use was categorized as having "high" influence on budget allocations (Global Fund, GAVI, USAID, DFATD). For example, for USAID a review found that more than one-third of evaluations led to budgetary changes (USAID, 2013f). The Global Fund uses evaluations to inform its Performance-Based Funding (PBF) model (The Global Fund, n.d)

Two organizations were classified as "medium" (WBG, DFAT), and one organization was categorized as "low" (Norad), due to pressure from within the organizations to disburse funds "without consideration of results

frameworks" (Grasso, Imas, & Fostvedt, 2012). For the remaining 11 donors, the evidence we review does not specify the extent to which results information influences budget allocation. Four organizations (Danida, WHO, WFP, AFD) describe plans to use results-based funding, but do not demonstrate operationalization of these plans in the documents we review.

4.7.3. Challenges to the Use of Results Information

Each donor organization faces challenges using results information. These challenges vary from simple issues involving a few key players or departments, to complex problems requiring changes at multiple levels of management and institutional cultural change. This section describes the three most reported challenges to using results information: reporting failures, creating quality usable evaluations, and a lack of a culture that values learning from evaluations.

Organizations reporting results have an incentive to withhold **negative results and failures**. We rate donors as "high" when there is evidence for specific processes and explicit cases for reporting and incorporating lessons from failures, as "medium" when there is some evidence of reporting failures, but no evidence for a specific process for learning, and as "low" when there are noted hindrances to reporting failure in the organization (column four, *Figure 12*). Five of the 22 we review have "high" attention to processes for recording failures (WBG, IMF, DFAT, DFATD, and Norad). We rated seven organizations as "medium" (AFD, JICA, GIZ, Sida, ADB, IADB, WFP), and three organizations (Danida, USAID, DFID) as "low," due to organizational hindrances to reporting. For example, DFID staff report feeling pressure to be positive (ICAI, 2014) and a review of USAID evaluations reports that unplanned results or alternative causes are not commonly recorded due to inadequate expertise on the evaluation teams (Hageboeck, Frumkin, & Monschein, 2013).

Several donor organizations report that recommendations or lessons learned from evaluations are not used by management or other authorizers because of **issues with the quality and timeliness of reports**. Frequently mentioned challenges include that evaluation reports are too long and difficult to read (Danida, USAID, Norad), and that they contain recommendations that are not specific enough to act on (Norad, ADB, WFP). One USAID review states that recommendations may not be used because recommendations are not connected to data (Blue, et al., 2009). An AfDB review report that results information is not used because of a long lag time (up to three years) between the completion of the project and the availability of results information (CIDA, 2013).

Multiple donor organizations also reported some element of a **weak culture for the use of results information** as a challenge. Danida reports a lack of demand by the Ministry of Foreign Affairs for the use of results information (InVirke & Publikum, 2014). Norad reports that a lack of an incentives system or sanctions hinders the use of results information and recommended that the Ministry of Foreign Affairs publish examples of instances when results information is used well to incentivize greater use of the information (Grasso, et al., 2012). An absent or weak culture of using results information may also be an underlying cause of DFAT's challenge with inconsistent instructions from management on how to describe the performance of programs (DFAT, 2014a; DFAT, 2014b). DFIT reports that the UN Group's "weak culture of striving for results" and the lack of support from management as well as a lack of buy-in from the organization at large has prevented the proper use of results information (DFID, 2011a).

5.   Trends in Donor Results Measurement Systems

Following our review of donor organizations' results measurement systems, we compare the systems used by different categories of donors by examining the extent to which different donors expend resources at varying levels among the different types of results measures. Not surprisingly (because of varying strategies and budget constraints), organizations give inconsistent attention to different types of results measurement , as organizations may for example have a strong focus on measuring outcomes but pay less attention to outputs of cost-effectiveness. We also find several distinguishing trends among groups of donor organizations.

Table 1 summarizes our ratings of donors' self-reported attention to different types of results measurement. Cases where we report "Not specified" indicate that donors do not provide information in public documentation of a particular type of results measure. For example, we rate GIZ in all types of evaluation because they provide this information in public documentation, whereas Danida has two types of results coded as "Not specified" because they do not provide information on these types of results measurement. Our rating on attention to spending is based on emphasis on spending in results reporting. We rate coordination with recipients based on the reporting and performance on the OECD's indicators of aid effectiveness. We rate outputs based on whether organizations report on outputs, compare outputs against targets, and specify output indicators. Similarly, the outcomes rating is based on whether organizations report on outcomes, compare

outcomes against targets, baseline, or control groups, and specify outcome indicators. We rate efficiency and cost-effectiveness based on organizations' reporting on comparing costs against outputs and outcomes, respectively.

Table 1: Attention to Different Types of Results Measurement

Organization	Region	Bilateral or Multilateral	Spending	Coordination with Recipients	Outputs	Outcomes	Efficiency: Comparing Outputs to Costs	Cost-Effectiveness: Comparing Outcomes to Costs
AFD	Europe	Bilateral	Medium	Medium-low	Medium-low	Medium	Not specified	Not specified
ADC	Europe	Bilateral	Not specified	Medium-low	Medium	Not specified	Not specified	Not specified
Danida	Scandinavia	Bilateral	Medium	High	Medium	Medium-low	Not specified	Not specified
DFAT	Asia	Bilateral	High	Medium	Medium-low	Medium-high	Low	Low
DFATD	Americas	Bilateral	High	Medium-low	Medium-low	Medium-high	Not specified	Low
DFID	Europe	Bilateral	High	Medium	High	Medium	High	Low
Dutch MFA	Europe	Bilateral	Not specified	Medium-high	Not specified	Not specified	Not specified	Not specified
GIZ	Europe	Bilateral	High	Medium-high	Medium-low	Medium	High	Medium
JICA	Asia	Bilateral	Low	Medium-high	Medium-low	Low	Low	Low
Norad	Scandinavia	Bilateral	Medium	Medium-low	Low	Low	Low	Not specified
SIDA	Scandinavia	Bilateral	Medium	High	Medium-high	Medium-low	Not specified	Not specified
USAID	Americas	Bilateral	Medium	Low	Medium-low	High	Medium	Low
AfDB	Africa	Multilateral	Medium	Medium-low	High	Medium	Medium	Not specified
ADB	Asia	Multilateral	High	Medium	High	High	Not specified	Not specified
GAVI Alliance	N.A.	Multilateral	High	Medium	Medium	High	Not specified	High
Global Fund	N.A.	Multilateral	High	Medium-low	High	Medium	High	Medium
IADB	Americas	Multilateral	High	Medium	High	High	Medium	High
IMF	N.A.	Multilateral	Medium	Medium-high	Medium-low	Low	Not specified	Not specified
UNG	N.A.	Multilateral	Medium	Medium-low	Low	Medium-low	Not specified	Not specified
WHO	N.A.	Multilateral	High	Not specified	Medium-high	Medium-low	Not specified	Not specified
WBG	N.A.	Multilateral	Medium	Medium-high	High	High	Not specified	High
WFP	N.A.	Multilateral	High	Not specified	High	High	Low	Low

While it is difficult to generalize a taxonomy of results measurement from our findings, we do identify some general trends. In this section, we consider differences in results measurement between multilateral and bilateral donors and between donors based in different regions of the world, and report other trends in results measurement.

5.1 Differences Between Multilateral and Bilateral Donors

Our sample includes 12 bilateral donors (Danida, Norad, Sida, AFD, ADC, DFID, MFA, GIZ, DFAT, JICA, DFATD, USAID) and ten multilateral donors (GAVI, Global Fund, IMF, UNG, WHO, WBG, WFP, ADB, IADB, AfDB). Multilateral and bilateral donors

are similar in many categories, but we find differences in how their evaluation and results measurements systems are organized and in their attention to harmonization and coordination with recipients, outputs and outcomes, and costs.

Multilateral donors use independent, but internal, evaluation committees to conduct evaluations more than bilateral donors (five of ten multilateral donors; one of 12 bilateral donors), whereas bilateral donors report using external evaluators more often (one of ten multilateral donors; nine of 12 bilateral donors). Both multilateral and bilateral donors often require grantees to develop some type of theory for the causal model of the intervention to support alignment of grantee-level results measurement with donor results measurement. 80 percent of multilateral donors currently require some type of logic model, compared to just over 40 percent of bilateral donors , though two other bilateral donors plan to start requiring more developed intervention theories. Multilateral and bilateral donors also substantially differ in guidance or standards provided to grantees for collecting or aggregating data and verifying or assuring data quality (six of ten multilateral donors; four of 12 bilateral donors for each category).

Attention to coordination and alignment between donors and recipient countries is slightly greater among bilateral donors than multilateral donors. Just under half of bilateral donors receive a "medium-high" or better rating for coordination with recipient country systems based on the OECD Paris Declaration aid effectiveness indicators, while only two multilateral donors receive a rating better than "medium." Multilateral donors are more likely to meet targets for reducing the number of parallel project implementation units (PIUs) (four of ten multilateral donors; three of 12 bilateral donors), but bilateral donors are more likely to meet targets for coordinating donor missions (one of ten multilateral donors; five of 12 bilateral donors). All donor organizations, however, perform poorly for coordinating recipient-country analyses (one of ten multilateral donors; two of 12 bilateral donors).

Multilateral donors' attention to outputs and outcomes stands out compared to what is reported by bilateral donors (*Figure 15*). Over half of multilateral donors receive a "high" rating for attention to outputs based on whether they report outputs, compare outputs against targets, and specify particular output indicators. The same proportion of multilaterals also receive a "high" rating for attention to outcomes based on whether they report outcomes, specify outcome indicators, and compare outcomes against targets, a baseline, or control groups. Multilateral donors more often report comparing output results to targets to assess performance (seven of ten multilateral donors; two of 12 bilateral donors), compare outputs across projects (four of ten multilateral donors; zero of 12 bilateral donors), specify collecting particular output indicators (seven of ten multilateral donors; four of 12 bilateral donors), and disaggregate output indicators by gender (six of ten multilateral donors; one of 12 bilateral donors). The differences between multilateral and bilateral donors in terms of attention to outcomes are less stark, as the share of bilateral and multilateral donors reporting on outcomes is about even. Multilateral donors are, however, more likely to demonstrate the connection between the outcomes and the theoretical framework (results framework, logic model, theory of change, or similar) in outcome or impact evaluations (six of ten multilateral donors; three[26] of 12 bilateral donors).

We did not find any differences between bilateral and multilateral donors in their attention to reporting spending results, or in comparing costs to outputs to measure implementation efficiency. We do, however, find that multilateral donors are more likely to compare costs to outcomes in order to evaluate cost-effectiveness; all three donors rated as "high" attention to cost-effectiveness are multilaterals (GAVI, IADB, WBG). Just one bilateral donor (GIZ) is even rated "medium," while five describe minimal attention to cost-effectiveness and six make no mention of comparing costs to outcomes.

---

[26] Three additional bilateral donors describe a connection to its theoretical base in impact evaluations for some of their programs, but not all (USAID, JICA, Norad) and two plan to require that the connection between the outputs and intended impact be made clear (Global Fund, UN Group).

Figure 15: Differences between Multilateral and Bilateral Donor Organizations

Reports outputs | Compares outputs against targets | Specifies output indicators | Requires outputs to be disaggregated by gender

- Multilateral
- Bilateral

Reports outcomes | Compares outcomes against targets | Specifies outcome indicators | Connects evaluations to theory of change

- Multilateral
- Bilateral
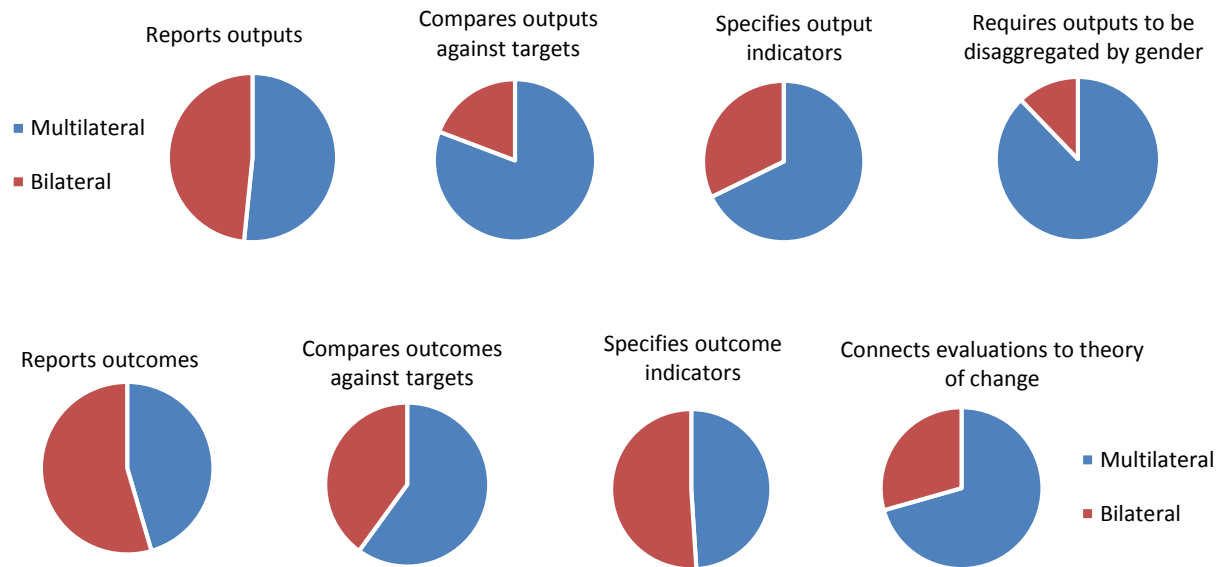
5.2 Differences Among Regions of Donors

We also find some associations between donor regions and attention to particular results though the sample size of any given region is small. Our sample includes donors from Africa, Asia and the Pacific, the Americas, and Europe, though we consider the three Scandinavian bilateral donors separately because these donors are often regarded as the most effective among bilateral donors. The donor regions, which most multilateral donors[27] do not fit into, break down as follows: one from Africa (AfDB), three from the Americas (DFATD, USAID, IADB), three from the Asia and Pacific region (DFAT, JICA, ADB), five from Europe (AFD, ADC, DFID, MFA, GIZ), and three from Scandinavia (Danida, Norad, Sida). Most regional differences are slight due to the small sample of any given region, the largest being Europe with five donors and the smallest being Africa with only the African Development Bank. Scandinavian donors stand out for their harmonization with recipient country systems and donors from the Americas give the most attention to outcomes and impacts.

We find that donors from Africa, Asia and Pacific, and the Americas report a process for reviewing and evaluating at the grantee and country level in addition to a process for reviewing and evaluating aggregated performance across all grants. Scandinavian donors are most harmonized with country-recipients and donors from the Americas are least harmonized. Based on the OECD Paris Declaration aid effectiveness indicators and peer review, Scandinavian donors receive the highest overall rating for donor harmonization with recipient country systems (two of three rate "high," one rates "medium-low"). Donors from the Americas have the lowest levels of harmonization (one rates "low," one "medium-low," and one "medium").

On the other hand, American donors receive the best overall rating for their attention to outcomes and impact and Scandinavian donors receive the lowest. Donors from the Americas report outcomes and specify particular outcome indicators, and compare outcomes against targets, baselines, or control groups more than any other region (we rate two of three "high" and one "medium-high"). Scandinavian donors receive the lowest rating for this category (two rating "medium-low," and one rating "low"). This finding indicates that attention to the OECD's measures of aid effectiveness may direct attention away from attention to other measures of organizational performance.

Donors from the Asia and Pacific region and Europe report their spending more than other regions (Asia and Pacific region two of three receive a "high" rating, of the three European donors that report addressing this issue, two of them receive a "high" rating, three "not specified"). Asia and Pacific region donors do least to compare costs against outputs or outcomes

---

[27] Multilateral donors for whom a region was not assigned include: GAVI Alliance, Global Fund, International Monetary Fund, UN Group, World Health Organization, World Bank Group, and World Food Program.

(for both comparing against outputs and outcomes two of the three Asia and Pacific donors rate "low"). While Europe is the only region with a "high" rating for comparing spending against outputs (the only two donors who report were rated "high", three others were "not specified"), these same two donors rate poorly for comparing spending against outcomes (one rated "low" the other rated "medium" and three were "not specified"). The IADB in the Americas region is the only donor to receive a "high" rating for comparing spending against outcomes, whereas the other two donors in the Americas region rate "low."

5.3 Other Trends in Results Measurement

Based on the ratings of attention to different areas of results measurement (presented in *Table 1*), we can identify which donors consider the most measures of results. Eight organizations (DFID, GIZ, ADB, GAVI, Global Fund, IADB, WBG, WFP) have strong attention to various measures of results, with "medium-high" or "high" ratings in three or more areas. These eight donors include two bilateral and six multilateral donors. While the areas these organizations report on vary, most emphasize reporting on spending (seven of eight), outputs (six), and outcomes (five), while few emphasize coordination with recipients (two). It is not clear what characteristics these eight organizations have in common that may contribute to their greater attention to multiple types of results measurement.

Four donor organizations (ADC, MFA, JICA, Norad) receive a "medium-high" or "high" rating in one or fewer areas of results measurement. For DAC and MFA, this is largely because the documents we reviewed do not provide any information on most areas of results measurement. Both JICA and Norad, however, have low attention to most areas of results measurement.

Overall, donors are more likely to consider costs by amount spent than by efficiency or cost-effectiveness. 19 of 22 donors are rated "medium" or "high" for attention to reporting spending results, but just six have similar ratings for comparing costs to outputs to measure efficiency, while five have similar ratings for comparing costs to outcomes to measure cost-effectiveness. Interestingly, the three donors with "high" attention to efficiency (DFID, GIZ, Global Fund) differ from the three with "high" attention to cost-effectiveness (IADB, GAVI Alliance, WBG).

We also find that donor organizations that focus on aid coordination and the OECD's indicators of aid effectiveness appear less likely to measure other types of results. Of the seven organizations (Danida, Sida, JICA, GIZ, WBG, IMF, Dutch MFA), just two (GIZ, WBG) receive "high" ratings for attention to another area of results measurement. It is not the case, however, that donors not focusing on aid coordination results have higher attention to other areas of results measurement. Of the eight donors rated "medium-low" or "low" in this area, four have "high" attention to at least one other measure of results, but four do not have any other ratings higher than "medium."

6. **Conclusion**

Donors want aid allocations to provide intended effects for the targeted demographics and to maximize impact, but donors may assess achievements differently depending on the sector and indicators of interest. Donor organizations and agencies can use results information for external accountability, institutional accountability and strategic management, and management and improvement of programs, projects, and policy (OECD, 2013c). They achieve these objectives by evaluating coordination between donors and recipients and the outputs, outcomes, costs, efficiency, and cost-effectiveness of funded interventions. Approaches to evaluation based on results-based management (DAC, 2000) and value for money (DFID, 2011c) are becoming increasingly important throughout the sector, and the Paris Declaration on Aid Effectiveness and the Accra Agenda for Action in 2008 play an important role in shaping the results measurement priorities of bilateral and multilateral donor organizations (Porter & Goldman, 2013; Segone 2010). Standard methods for results measurement have not been established across donor organizations, however, leading to significant variability in donor's results measurement systems. This report captures the variation in donor results measurement systems, including how these systems are organized, the levels of results they report, and the types of results information they measure. We review peer evaluations and donor documentation describing 22 donor evaluation and results measurements systems, including systems for 12 bilateral organizations and ten multilateral organizations.

Donor organizations usually have established processes to measure and evaluate results. The levels at which donors evaluate results vary, as does the amount of guidance or technical support the donor organization offers for evaluation. Regular evaluations at the aggregate donor level (summarizing performance across projects, countries, or portfolios) are common, occurring every one to three years. Evaluations at the grantee or project level are supported by donor organizations through guidelines and technical assistance. Most donor organizations evaluate at more than one level, with overarching evaluation performed at the donor level based on aggregated evaluations from at least one other lower level

(grantee/project, country, portfolio). Coordination between what donors want to measure at the aggregate donor-level and what they measure and evaluate at lower levels is an important factor in donor's ability to evaluate the impact of their aid.

A key element in coordinating evaluations is defining what results to measure or track. While most donors measure and report results in terms of outputs and outcomes, many do not have donor-level systems in place defining what indicators to monitor, instead determining indicators at a lower level, which can create challenges with data quality and comparisons across projects and over time. As a result, reporting of output and outcome results is inconsistent and donors face difficulties aggregating these individual project results. In spite of these challenges, over three quarters of donors (17 of 22) report output results and over half (12 of 22) report outcome results at the donor level. Multilateral donor organizations are, however, more likely to specify output and outcome indicators at the donor level and report systems in place for collecting and aggregating those data from individual projects and partners up to the donor level.

Costs and spending is another area of measurement that donors use to evaluate performance. While reporting aid allocations and spending is a common measure of donor activity, few organizations use spending measures in combination with measures of outputs or outcomes to evaluate implementation efficiency or cost-effectiveness of interventions.

A final type of results measurement, prioritized by the Paris Declaration of Aid Effectiveness and other later agreements, is measuring aid effectiveness by coordination and alignment between donors and recipients. These measures are a prominent feature of reviews of aid effectiveness and peer reviews published by the OECD. Many donors, especially bilateral donors, describe taking action towards improving their alignment and harmonization with country-level systems. Challenges remain, however, as donors face difficulties in coordinating with diverse stakeholders and coordinating joint evaluations.

We do not find clear or consistent differences in approaches to results measurement among donor regions or between bilateral and multilateral donors, though some trends distinguish these groups of donors. Scandinavian donors are rated highest in our review for coordinating with recipient-countries, and donors based in the Americas are rated higher than other regions for their attention to outcome and impact measurement. Bilateral donors rate higher on measures of coordination with recipient-country systems, while multilateral donors tend to be more explicit about the attention they give to measuring outputs and outcomes and are more likely to compare costs to outcomes to evaluate cost-effectiveness.

Our review identifies many challenges donor organizations face and recommendations for addressing these issues. Across the types of results measurement we review, resources and coordination for results measurement are the main limitations described by donors. Resource limitations in terms of funding, budgets, staff, training, and expertise constrain donors' ability to carry out results measurement as they intend. Inadequate resources for data collection also affect the availability of baseline data and data quality, and thus create challenges for overall results measurement and applicability of results information to decision-making. Coordination challenges are more complex and vary across donors. Some donors express difficulty identifying and providing the necessary guidelines or technical support needed at each level of results measurement, while others face challenges aligning measures with diverse stakeholders.

Recommendations for how to address these challenges are often donor specific; however, some are more common. Specific donor recommendations include changing the organizational culture around evaluation, clarifying policies for data validation, and using real-time monitoring systems as methods to improve data quality, collection, and reporting across different levels of evaluation. For improving coordination across levels of results measurement, donors report success with shared dissemination of results, systematic feedback loops between donors and countries, and streamlined processes.

While the strength of evidence in our review is mostly "high" (as illustrated in *Figure B.1* in Appendix B), there are some gaps in the evidence we review. For over half of our sample, the evidence we reviewed did not include information on whether donors compare costs to outputs and whether they compare costs to outcomes. Even fewer donors report on whether or not they compared costs against budgets. Reviewing budget or audit documentation for these donors could provide additional information on these areas of results measurement. Further, we find little information in the evidence we reviewed on donors' practices of comparing output or outcome results across projects, collecting specified output indicators, and comparing outcomes against a control group.

An important difficulty in this review was identifying appropriate aggregate organization-level documentation on results measurement to review. As we aimed to review trends in donors' approaches to aggregate results measurement across all projects, countries, or portfolios, we did not review documents that only described these lower levels of results measurement. Since donors describe some difficulties and challenges with aggregated results and with coordinating their systems for results measurement, our review may not fully capture the extent of results measurement conducted by

donors. Further review of the differences between what results donors measure and report at different levels would add to our analysis of which areas of results measurement are lacking at the donor level, and what the key challenges donors face in measuring and aggregating results information are.

Please direct comments or questions about this research to Leigh Anderson, at eparx@u.washington.edu.

References Cited in Text

African Development Bank Group. (2013c). The One Bank Results Measurement Framework 2013-2016.
<http://www.afdb.org/fileadmin/uploads/afdb/Documents/Project-and-
Operations/The_One_Bank_Result_Measurement_Framework_2013-2016.pdf>.

African Development Bank. (2013d). Work Programme proposals (2014-2016) for the Independent Evaluation Department of the African Development Bank. <http://www.ecgnet.org/document/afdb-idev-work-programme-2014-2016>.

African Development Bank Group. (2014b). Annual Development Effectiveness review 2014.
<http://www.afdb.org/fileadmin/uploads/afdb/Documents/Project-and-Operations/ADER-
_Annual_Development_Effectiveness_Review_2014.pdf>.

AFD. (2013). AFD's Evaluation Policy.
<http://www.afd.fr/webdav/site/afd/shared/RECHERCHE/Evaluation/document/AFD_evaluation_policy.pdf>.

Asian Development Bank. (2013a). 2013 Development Effectiveness Review.
<http://www.adb.org/sites/default/files/institutional-document/42747/files/defr-2013-report.pdf>.

Asian Development Bank. (2013c). ADB Independent Evaluation Operations Manual 2013.
<http://www.adb.org/sites/default/files/institutional-document/31483/omk1.pdf>.

Asian Development Bank. (2014c). ADB's Results Framework: Interim Update to Align with the Midterm Review of Strategy 2020. <http://www.adb.org/sites/default/files/institutional-document/153880/r182-14.pdf>.

Asian Development Bank Independent Evaluation Department. (2015). Annual Evaluation Review 2015.
<http://www.adb.org/sites/default/files/AER-2015.pdf>

Binnendijk, A. (2000). Results Based Management in the Development Co-operation Agencies: A Review of Experience – Background Report. DAC Working Party on Aid Evaluation.
<http://www.oecd.org/development/evaluation/1886527.pdf>.

Blue, R., Clapp-Wincek, C., & Benner, H. (2009). Beyond Success Stories: Monitoring & Evaluation for Foreign Assistance Results: Evaluator Views of Current Practice and Recommendations for Change.
<http://www.usaid.gov/sites/default/files/documents/1868/060909_beyond_success_stories.pdf>.

Canadian Chief Audit Executive. (2012b). Internal Audit of Results-Based Management: Final Report.
<http://www.international.gc.ca/department-ministere/audits-verification/2012/dev-rbm-gar.aspx?lang=eng>.

CIDA. (2013). Development Effectiveness Review of the World Health Organization (WHO) 2007-2010.
<http://www.oecd.org/derec/18_WHO_DE_REVIEW_DONOR_NEUTRAL_VERSION_FOR_DAC.pdf>.

Department for International Development (DFID). (2011a). Multilateral Aid Review: Ensuring Maximum Value for Money for UK Aid through Multilateral Organizations.
<https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/67583/multilateral_aid_review.pdf>.

Department for International Development (DFID). (2011c). DFID's Approach to Value for Money (VfM).
<https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/67479/DFID-approach-value-money.pdf>.

Department for International Development (DFID). (2013). Multilateral Aid Review Update.
<https://www.gov.uk/government/publications/multilateral-aid-review-update-2013>.

Department for International Development (DFID). (2014). DFID Evaluation Annual Report from 1 July 2013 to 30 June 2014.
<https://www.gov.uk/government/publications/dfid-evaluation-annual-report-from-1-july-2013-to-30-june-2014>.

Department of Foreign Affairs and Trade. (2014a). 2013 Quality Review of Aid Program Performance Reports.
<http://agencysearch.australia.gov.au/search/click.cgi?rank=1&collection=agencies&url=http%3A%2F%2Fwww.dfat.gov.au%2Faid%2Fhow-we-measure-performance%2Fode%2FDocuments%2F2013-quality-review->.

Department of Foreign Affairs and Trade. (2014b). Learning from Australian Aid Operational Evaluations. <http://www.dfat.gov.au/aid/how-we-measure-performance/ode/Documents/learning-from-aust-aid-op-evaluations.pdf>.

Development Assistance Committee (DAC). (2013). DAC Peer Review of Sweden. <http://www.oecd-ilibrary.org/development/oecd-development-co-operation-peer-reviews-sweden-2013/organigram-sida_9789264196254-17-en>.

Deutsche Gesellschaft für Internationale Zusammenarbeit (GIZ). (2012). Measuring - Assessing - Making Improvements. <http://www.giz.de/en/downloads/giz2013-en-evaluierungsbericht-2010-2012.pdf>.

DFAT Office of Development Effectiveness. (2013). Rolling Evaluation Work Plan: 2013–14 to 2015–16. <http://dfat.gov.au/aid/how-we-measure-performance/ode/Documents/ode-evaluation-workplan.pdf>.

DFAT Office of Development Effectiveness. (2014). Quality of Australian Aid Operational Evaluations. <http://www.dfat.gov.au/aid/how-we-measure-performance/ode/Documents/quality-of-aust-aid-op-evaluations.pdf>.

DFATD. (2014a). Departmental Performance Report 2013-14. <http://www.international.gc.ca/department-ministere/assets/pdfs/DFATD-DPR-2013-14_FINAL_ENG.pdf>.

Federal Ministry for Economic Cooperation and Development (BMZ). (2013). The German government's 14th Development Policy Report. Bonifatius Druck und Verlag: Paderborn, Germany.

GAVI Alliance. (2011). Monitoring and Evaluation Framework and Strategy. <http://www.gavi.org/library/documents/gavi-documents/strategy/gavi-alliance-monitoring-and-evaluation-framework-and-strategy-2011-2015/>.

GAVI Alliance Board. (2012) GAVI Alliance Evaluation Advisory Committee Meeting 19-20 January 2012. GAVI. <http://www.gavi.org/About/Governance/GAVI-Board/Evaluation-Advisory-Committee/2012/Evaluation-Advisory-Committee-Meeting,-19-20-January-2012,-Final-Minutes/>.

GAVI Alliance. (2013b). GAVI Alliance Progress Report. <http://gaviprogressreport.org/2013/>.

The Global Fund. (n.d). Monitoring and Evaluation Framework Presentation Slide. <theglobalfund.org>.

Grasso, P.G., Imas, L.M., & Fostvedt, N. (2012). Use of Evaluations in the Norwegian Development Cooperation System. RAND Europe. <http://www.norad.no/globalassets/import-2162015-80434-am/www.norad.no-ny/filarkiv/evalueringsavdelingens-filer/use-of-evaluation-in-the-norwegain-development-cooperation-system.pdf>.

Hageboeck, Frumkin, & Monschein. (2013). META-EVALUATION OF QUALITY AND COVERAGE OF USAID EVALUATIONS 2009 – 2012. <http://usaidlearninglab.org/library/meta-evaluation-quality-and-coverage-usaid-evaluations-2009-2012>.

HLSP. (2009). GAVI Health Systems Strengthening Support Evaluation 2009. GAVI. <http://www.gavi.org/Library/Documents/GAVI-documents/Evaluations/GAVI-Health-Systems-Strengthening-support-evaluation-2009--Volume-2-full-evaluation-report/>.

IEG. (2012a). Comparison of the M&E Systems of the World Bank and the Global Fund. <http://ieg.worldbankgroup.org/Data/reports/gfr_wp_april2012.pdf>.

IEG. (2012b). Results and Performance of the World Bank Group 2013: An Independent Evaluation - Volume I: Main Report. <https://ieg.worldbankgroup.org/Data/reports/chapters/rap2013_vol1_updated2.pdf>.

IEG. (2014). Results and Performance of the World Bank Group 2013: An Independent Evaluation - Volume I: Main Report <https://ieg.worldbankgroup.org/Data/reports/chapters/rap2013_vol1_updated2.pdf>.

Independent Commission for Aid Impact (ICAI). (2012). DFID's Oversight of the EU's Aid to Low-Income Countries. <http://icai.independent.gov.uk/wp-content/uploads/2012/12/ICAI-EU-report-061212-FINAL.pdf>.

Independent Commission for Aid Impact (ICAI). (2014). How DFID Learns. <http://icai.independent.gov.uk/wp-content/uploads/2014/04/How-DFID-Learns-FINAL.pdf>.

Independent Development Evaluation. (2015). Independent Development Evaluation (IDEV) of the African Development Bank. <http://independentevaluation.afdb.org/en/>.

Independent Evaluation Office of the IMF. (2014e). IMF Technical Assistance: Revisiting the 2005 IEO Evaluation. <http://www.ieo-imf.org/ieo/files/updates/IEO%20-%20TA%20Mini%20Review%20w%20statement.pdf>.

Independent Evaluation Office of the IMF. (2015a). POSSIBLE TOPICS FOR EVALUATION OVER THE MEDIUM TERM. <http://www.ieo-imf.org/ieo/files/workprogram/Possible_Topics_for_Evaluation_Over_the_Medium_Term%202015%201.pdf>.

Institute for Health Metrics and Evaluation. (2013). Project Overview. <http://www.healthdata.org/gavi/project-overview>.

Inter-American Development Bank. (2014c). 2014 Annual Report. <http://publications.iadb.org/handle/11319/6855?locale-attribute=en>.

Inter-American Development Bank (2014b). Development Effectiveness Overview. <http://www.iadb.org/en/office-of-strategic-planning-and-development-effectiveness/development-effectiveness,1222.html>.

InVirke and Publikum. (2014). Danish MFA's evaluations of development assistance: Evaluations – use, user relevance and communication. <http://um.dk/en/~/media/UM/Danish-site/Documents/Danida/Resultater/Eval/Ref%20doc/Invirkerapport.pdf>.

IMF. (2014b). Annual Report 2014. <http://www.imf.org/external/pubs/ft/ar/2014/eng/pdf/ar14_eng.pdf>.

Itad. (2014). Can We Demonstrate the Difference that Norwegian Aid Makes? Evaluation of results measurement and how this can be improved. <http://www.oecd.org/derec/norway/Can-We-Demonstrate-the-Difference-that-Norwegian-Aid-Makes.pdf>.

Japan Ministry of Foreign Affairs. (2014). Review of Japan's ODA Evaluations from FY 2003 to 2013. <http://www.mofa.go.jp/policy/oda/evaluation/FY2014/pdfs/review_03-13.pdf>.

Japan Ministry of Foreign Affairs. (2014a). Annual Report on Japan's ODA Evaluation 2014. <http://www.mofa.go.jp/policy/oda/evaluation/annual_report_2014/index.html>.

JICA Evaluation Department. (2014b). JICA Guidelines for Operations Evaluation. <http://www.jica.go.jp/english/our_work/evaluation/tech_and_grant/guides/c8h0vm000001rfr5-att/guideline_2014.pdf>.

Kvalvaag, Tale & Stern, Elliott. (2014). A Peer Review of Danida's Evaluation Function. <http://um.dk/~/media/UM/Danish-site/Documents/Danida/Resultater/Eval/2014PeerReview.pdf>.

Lawson, Marian Leonardo. (2013). Does Foreign Aid Work? Efforts to Evaluate U.S. Foreign Assistance. *Congressional Research Service*. <http://www.fas.org/sgp/crs/row/R42827.pdf>.

Lloyd, R., Poate, D., & Villanger, E. (2014). Results measurement and evaluability: a comparative analysis. Journal of Development Effectiveness, 6(4), 378-391.

Macro International Inc. (2009). Global Fund 5-year Evaluation: Study Area 3: The Impact of Collective Efforts on the Reduction in Disease Burden. Retrieved from <theglobalfund.org>.

Ministry of Foreign Affairs of Denmark. (2011). Danish Development Cooperation in a Results Perspective: Danida's Framework for Managing for Development Results 2011-2014. <http://amg.um.dk/~/media/amg/Documents/Front%20page/DanidasFrameworkforManagingforDevelopmentResults20112014Final.pdf>.

Ministry of Foreign Affairs of Denmark. (2012b). Danida Evaluation Guidelines. <http://www.netpublikationer.dk/um/11121/index.htm>.

Norad. (2013). Annual Report 2013: Evaluation of Norwegian Development Cooperation. <http://www.norad.no/en/toolspublications/publications/2014/annual-report-2013-evaluation-of-norwegian-development-cooperation/>.

OECD. (2009). Austria: Development Assistance Committee (DAC) Peer Review. <http://www.oecd.org/dac/peer-reviews/austria%20full%20report.pdf>.

OECD. (2010). Quality Standards for Development Evaluation. <http://www.oecd.org/development/evaluation/qualitystandards.pdf>.

OECD. (2011). Aid Effectiveness 2011: Progress in Implementing the Paris Declaration. OECD Publishing.

OECD. (2013a). OECD Development Co-operation Peer Review: France 2013. <http://www.oecd.org/dac/peer-reviews/OECD%20France_ENGLISH%20version%20onlineFINAL.pdf>.

OECD. (2013c). Development Results - An Overview of Results Measurement and Management. <http://www.oecd.org/dac/peer-reviews/Development-Results-Note.pdf>.

OECD. (2014). United Kingdom - DAC Peer Review of Development Co-operation. <http://www.oecd.org/dac/peer-reviews/peer-review-unitedkingdom.htm>.

OECD. (2014b). Measuring and Managing Results in development co-operation. <http://www.oecd.org/dac/peer-reviews/Measuring-and-managing-results.pdf>.

OECD. (2014c). Statistics on Resource Flows to Developing Countries. <http://www.oecd.org/dac/stats/statisticsonresourceflowstodevelopingcountries.htm>.

OECD. (2014d). Aid to developing countries rebounds in 2013 to reach an all-time high. <http://www.oecd.org/newsroom/aid-to-developing-countries-rebounds-in-2013-to-reach-an-all-time-high.htm>.

Sinha, P. K. (2010) Can a Useful Aid Effectiveness Index Be Developed using the Paris Declaration Framework? <http://s3.amazonaws.com/aiddata/Sinha_aiddata.pdf>.

UNEG. (2014). Peer Review of the Evaluation Function of the UN World Food Programme (2008-2013). <http://documents.wfp.org/stellent/groups/public/documents/reports/wfp264679.pdf>.

United Nations Evaluation Group. (2010). UNEG Quality Checklist for Evaluation Reports. <http://www.uneval.org/document/download/853>.

USAID. (2013c). EVALUATION AT USAID NOVEMBER 2013 UPDATE. < http://usaidlearninglab.org/library/evaluation-usaid-november-2013-update>.

USAID. (2013e). AGENCY FINANCIAL REPORT FISCAL YEAR 2013. <http://www.usaid.gov/results-and-data/progress-data/agency-financial-report>.

USAID. (2013f). FY 2013 Joint Summary of Performance and Financial Information. <http://www.usaid.gov/results-and-data/progress-data/summary>.

Vivideconomics. (2011). Monitoring and evaluation frameworks and the performance and governance of international funds. <http://www.afdb.org/fileadmin/uploads/afdb/Documents/Generic-Documents/Monitoring%20and%20evaluation%20frameworks%20and%20the%20performance%20and%20governance%20of%20international%20funds.pdf>.

Wiig, H. & Holm-Hansen, J. (2014). Unintended Effects in Evaluations of Norwegian Aid: A desk study. Norwegian Institute of Urban and Regional Research (NIBR). <http://www.norad.no/globalassets/import-2162015-80434-am/www.norad.no-ny/filarkiv/vedlegg-til-publikasjoner/unintended-effects-in-evaluations-of-norwegian-aid.pdf>.

World Food Program (WFP). (2014). OPERATION EVALUATIONS Synthesis Report: July 2013/July 2014. <http://documents.wfp.org/stellent/groups/public/documents/reports/wfp268837.pdf>.

Other References Not Cited in Text but Used in Review

AFD. (2010). L'action de l'Agence Française de Développement pour les objectifs du Millénaire. <http://www.afd.fr/webdav/site/afd/shared/PUBLICATIONS/INSTITUTIONNEL/Action-developpement/AFD_OMD.pdf>.

AFD. (2012). Rapport conjoint au Parlement sur les évaluations de l'aide publique au développement française, 2010/2012. <http://www.afd.fr/webdav/site/afd/shared/RECHERCHE/Evaluation/document/Evaluation-Rapport-conjoint-parlement-afd-2010-2012.pdf>.

AFD. (2014a). AGENCE FRANÇAISE DE DÉVELOPPEMENT: ANNUAL REPORT 2013. <http://www.afd.fr/webdav/site/afd/shared/PUBLICATIONS/INSTITUTIONNEL/rapports-annuels/annual-report-afd-2013.pdf>.

AFD. (2014b). AGENCE FRANÇAISE DE DÉVELOPPEMENT: OVERVIEW. <http://www.afd.fr/webdav/site/afd/shared/PUBLICATIONS/INSTITUTIONNEL/plaquettes-presentation/AFD-bref-EN.pdf>.

AFD. (2015a). Family Farming. <http://www.afd.fr/webdav/site/afd/shared/PUBLICATIONS/THEMATIQUES/AFD-agriculture-familiale-VA.pdf>.

AFD. (2015b). Ghana. <http://www.afd.fr/webdav/site/afd/shared/PORTAILS/PAYS/GHANA/PDF/afd-ghana-va.pdf>.

AFD. (2015c). European Union. <http://www.afd.fr/webdav/site/afd/shared/PUBLICATIONS/THEMATIQUES/AFD-Union-europeenne-VA.pdf>.

AFD. (n.d.). INCREASING AND DEMONSTRATING AID EFFECTIVENESS. <http://www.afd.fr/webdav/site/afd/shared/PUBLICATIONS/INSTITUTIONNEL/Action-developpement/TriptiqueGRD_GB.pdf>.

African Development Bank. (2011). Evaluation of Paris Declaration Implementation at the African Development Bank. <http://www.afdb.org/en/documents/document/evaluation-of-paris-declaration-implementation-at-the-african-development-bank-25388/>.

African Development Bank. (2012). Fostering Regional Integration in Africa: An Evaluation of the Bank's Multinational Operations, 2000-2010. <http://www.oecd.org/derec/afdb/7_FosteringRegionalintergrationinAfricaMutinational%20Operations.pdf>.

African Development Bank. (2013a). At the Center of Africa's Transformation: Strategy for 2013–2022. <http://www.afdb.org/fileadmin/uploads/afdb/Documents/Policy-Documents/AfDB_Strategy_for_2013%E2%80%932022_-_At_the_Center_of_Africa%E2%80%99s_Transformation.pdf>.

African Development Bank. (2013b). Annual Report <http://www.afdb.org/fileadmin/uploads/afdb/Documents/Publications/Annual_Report_2013.pdf>.

African Development Bank. (2013e). Botswana Country Assistance Evaluation 2004-2013. <http://independentevaluation.afdb.org/en/evaluations-publications/evaluation/botswana-country-assistance-evaluation-2004-2013-511/>.

African Development Bank. (2014a). Congo Basin Forest Fund 2013 Annual Report. <http://www.afdb.org/fileadmin/uploads/afdb/Documents/Evaluation-Reports-_Shared-With-OPEV_/Annual_Report_%E2%80%93_Congo_-_Congo_Basin_Forest_Fund_2013_-_CBFF_%E2%80%93_10_2014.pdf>.

Armytage, Livingston. (2011). Evaluating aid: an adolescent domain of practice. *Sage*.

Asian Development Bank. (2011a). The Asian Development Fund Operations: A Decade of Supporting Poverty Reduction in the Asia and Pacific Region. <http://www.oecd.org/derec/adb/49049850.pdf>.

Asian Development Bank. (2011b). Managing for Development Results. <http://www.oecd.org/derec/adb/49049141.pdf>.

Asian Development Bank. (2013b). Work Program and Budget Framework 2014-2016. <http://www.adb.org/sites/default/files/institutional-document/33983/files/work-program-budget-framework-2014-2016.pdf>.

Asian Development Bank. (2014a). 2014 Annual Evaluation Review. <http://www.adb.org/sites/default/files/evaluation-document/111009/files/2014-aer.pdf>.

Asian Development Bank. (2014b). Midterm Review of Strategy 2020: Meeting the Challenges of a Transforming Asia and Pacific. <http://www.adb.org/sites/default/files/institutional-document/34149/files/midterm-review-strategy-2020-r-paper.pdf>.

Asian Development Bank. (2014d). Operations Manual Bank Policies: Performance-based Allocation of Asian Development Fund Resources. <http://www.adb.org/sites/default/files/institutional-document/31483/om-a3.pdf>.

Asian Development Bank Independent Evaluation Department. (2015). Annual Evaluation Review 2015. <http://www.adb.org/sites/default/files/AER-2015.pdf>.

AusAID Office of Development Effectiveness. (2010). Annual Review of Development Effectiveness 2009. <http://www.oecd.org/derec/australia/46755540.pdf>.

AusAID. (2013). AusAID Annual Report 2012-13. <http://www.oecd.org/derec/australia/Ausaid-annual-report-2012 13.pdf>.

Australian Government. (2011). Independent Review of Aid Effectiveness. <http://www.aidreview.gov.au/publications/aidreview.pdf>.

Australian Government. (2012). Australian Multilateral Assessment March 2012 – GAVI Alliance. < http://dfat.gov.au/about-us/publications/Documents/gavi-assessment.pdf>.

Austrian Development Cooperation. (2008). GUIDELINES FOR EVALUATION IN AUSTRIA'S OFFICIAL DEVELOPMENT COOPERATION. <http://www.oecd.org/dataoecd/2/54/35126441.pdf>.

Austrian Development Agency Evaluation Unit. (2009b). "Guidelines for Project and Programme Evaluations. <http://www.entwicklung.at/uploads/media/Guidelines__ProjectProgamme_Evaluations_Juli2009_01.pdf>.

Bester, A. (2012). Results-Based Management in the United Nations Development System: Progress and Challenges. <http://www.un.org/esa/coordination/pdf/rbm_report_10_july.pdf>.

BMZ. (2009). Evaluation in German Development Cooperation: A System's Review. <http://www.oecd.org/derec/germany/48395320.pdf>.

BMZ. (2011). Tools and Methods for Evaluating the Efficiency of Development Interventions. <http://www.oecd.org/derec/germany/47796853.pdf>.

Cafferini, L. & Pierrel, H. (2009). Sectoral aid in practice. <http://www.afd.fr/webdav/site/afd/shared/PUBLICATIONS/RECHERCHE/Evaluations/Notes-synthese/07-VA-Notes-synthese.pdf>.

Canadian Chief Audit Executive. (2012a). Internal Audit of Grants and Contributions in Geographic Programs Branch. <http://www.international.gc.ca/department-ministere/audits-verification/2012/dev-gpb-dgpg.aspx?lang=eng>.

Canadian Internal Audit Division. (2005). Data Integrity Control Framework. <http://www.acdi-cida.gc.ca/INET/IMAGES.NSF/vLUImages/Internal%20Audits/$file/Data%20Integrity%20Control%20Framework.pdf>.

Carr, B. (2013). Budget: Australia's Development Assistance Program 2013-14. <http://www.budget.gov.au/2013-14/content/ministerial_statements/download/AusAid_MS.pdf>.

CEPA, LLP. (2010). Second GAVI Evaluation.
<http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.392.7981&rep=rep1&type=pdf>.

Chandy, L. (2011). Benchmarking against progress: An assessment of Australia's aid effectiveness. *Global Economy & Development, Working Paper* (46).

CIDA. (2006a). A Review of Evidence of the Effectiveness of CIDA's Grants and Contributions. <http://www.acdi-cida.gc.ca/acdi-cida/acdi-cida.nsf/eng/NAT-4109391-HZY>.

CIDA. (2006b). Peer Review of Evaluation Function at UNICEF. <http://www.uneval.org/document/detail/98>.

CIDA Evaluation Directorate. (2011a). Rolling Five-Year Evaluation Work Plan FY 2011/2012 to FY 2015/2016.
<http://www.acdi-cida.gc.ca/inet/images.nsf/vLUImages/Evaluations2/$file/five-year-rolling-work-plan-master-document-eng.pdf>.

CIDA Evaluation Directorate. (2011b). A Review of Evidence of the Effectiveness of CIDA's Grants and Contributions 2005/06 – 2010/11. <http://www.oecd.org/derec/canada/50452019.pdf>.

Clarke, J., & Visser-Valfrey, M. (2010). Results and Value for Money: A Performance Review of the Human Development Portfolio in Mozambique.
<https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/67676/perf-rev-hum-dvmt-portf-mz.pdf>.

Danida. (2005). Peer Assessment of Evaluation in Multilateral Organisations: United Nations Development Programme.
<http://www.uneval.org/document/detail/99>.

Department for International Development. (2005). Guidance on Evaluation and Review for DFID Staff.
<https://www.gov.uk/government/publications/guidance-on-evaluation-and-review-for-dfid-staff-guidance-evaluation>.

Department for International Development. (2011a). Multilateral Aid Review: Ensuring Maximum Value for Money for UK Aid through Multilateral Organizations.
<https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/67583/multilateral_aid_review.pdf>.

Department for International Development. (2011b). Multilateral Aid Review: Taking Forward the Findings of the UK Multilateral Aid Review.
<https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/200094/MAR-taking-forward.pdf>.

Department for International Development (2012). BROADENING THE RANGE OF DESIGNS AND METHODS FOR IMPACT EVALUATIONS. <http://www.dfid.gov.uk/Documents/publications1/design-method-impact-eval.pdf>.

Department for International Development. (2014b). DFID's Results Framework: Managing and reporting DFID results.
<https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/360906/DFID-external-results-Sep_2014.pdf>.

Department for International Development. (2014c). Annual Report and Accounts 2013–14.
<https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/331591/annual-report-accounts-2013-14a.pdf>.

Department of Foreign Affairs and Trade. (2014). Australia's New Development Policy and Performance Framework: A Summary. <www.dfat.gov.au/about-us/publications/Documents/aid-policy-summary-doc.docx+&cd=12&hl=en&ct=clnk&gl=us>.

Department of Foreign Affairs and Trade. (2014). Making Performance Count: Enhancing the Accountability and Effectiveness of Australian Aid. <www.dfat.gov.au/about-us/publications/Documents/framework-making-performance-count.docx+&cd=13&hl=en&ct=clnk&gl=us>.

Department of Foreign Affairs and Trade. (2015). Performance of Australian Aid: 2013-14. <http://dfat.gov.au/about-us/publications/Documents/performance-of-australian-aid-2013-14.pdf>.

Deutsche Gesellschaft fuer Internationale Zusammenarbeit GmbH. (2011). Methods and Instruments: Evaluation. <http://www.giz.de/en/downloads/giz2011-en-evaluation.pdf>.

Deutsche Gesellschaft fuer Internationale Zusammenarbeit GmbH. (2011). Methods and Instruments: Results-based monitoring. <http://www.giz.de/en/downloads/giz2011-en-result-based-monitoring.pdf>.

Deutsche Gesellschaft fuer Internationale Zusammenarbeit GmbH. (2011). Methods and Instruments: Social Impact Assessment. <http://www.giz.de/en/downloads/giz2011-en-social-impact-assessment.pdf>.

Development Assistance Committee. (2009). DAC Peer Review of Sweden. <http://www.oecd.org/development/peer-reviews/43278517.pdf>.

DFAT Office of Development Effectiveness. (2015). Research for better aid: an evaluation of DFAT's investments. <http://dfat.gov.au/aid/how-we-measure-performance/ode/Documents/research-for-better-aid-an-evaluation-of-dfats-investments.pdf>.

DFATD. (2013). Lessons from Development Evaluations – 2013. <http://www.international.gc.ca/department-ministere/evaluation/workplans_lessons-lecons_plans_travail/dev-lfde-lted13.aspx?lang=eng>.

DFATD. (2014b). Statistical Report on Canadian International Assistance: Fiscal Year 2013-2014. <http://www.international.gc.ca/development-developpement/assets/pdfs/sria-rsai-2013-14-eng.pdf>.

DFATD. (2014c). Report to Parliament on the Government of Canada's Official Development Assistance. <http://www.international.gc.ca/development-developpement/assets/pdfs/2013-14oda_report.pdf>.

DFATD. (2014d). Canada's Development Program at a Glance: 2013-2014. <http://www.international.gc.ca/development-developpement/assets/pdfs/15-015_D4R-light-E.pdf>.

DFATD. (2014e). Rolling Five-Year Development Evaluation Work Plan 2014-2015 - 2018-2019. <http://www.international.gc.ca/department-ministere/evaluation/workplans_lessons-lecons_plans_travail/dev-rfydewp-ptqcmed14.aspx?lang=eng>.

DFATD. (2014f). Report on Plans and Priorities 2015-2016. <http://www.international.gc.ca/department-ministere/assets/pdfs/RPP_2015_2016_ENG.pdf>.

FAO. (2012). Peer Review: The Evaluation Function of the Food and Agriculture Organization. <http://www.uneval.org/document/detail/1191>.

Federal Ministry for Economic Cooperation and Development (BMZ). (2015). Strategy on government-civil society cooperation in post-2015 development policy. Berlin: Germany.

Gajo, Michael. (2013). Lessons learnt from impact evaluation of GIZ interventions. Gesellschaft fuer Internationale Zusammenarbeit GmbH. <http://www.kdi.re.kr/data/download/attach/9831_2-3.pdf>.

GAVI Alliance Board. (2009). GAVI Evaluation. Retrieved from <gavi.org>.

GAVI Alliance Board. (2013) GAVI Alliance Evaluation Advisory Committee Meeting 11 July 2013. GAVI. Retreived on April 22, 2015 from <http://www.gavi.org/About/Governance/GAVI-Board/Evaluation-Advisory-Committee/2013/Evaluation-Advisory-Committee-Meeting,-11-July-2013,-Final-Minutes/>/

GAVI Alliance Board. (2014) GAVI Alliance Evaluation Advisory Committee. GAVI. Retrieved on April 22, 2015 <http://www.gavi.org/About/Governance/GAVI-Board/Governance-committee/2014/Governance-Committee-Meeting,-24-March-2014,-Final-Minutes/>

GAVI Secretariate. (2010). Report from Chair of Evaluation Advisory Committee to the Board. GAVI. Retrieved April 22, 2015 from <http://www.gavi.org/About/Governance/GAVI-Board/Minutes/2010/30-November/Minutes/Report-of-the-Evaluation-Advisory-Committee-Chair/>.

Government Accountability Office. (2012). Public Financial Management: Improvements Needed in USAID's and Treasury's Monitoring and Evaluation Efforts. <http://www.gao.gov/assets/650/648221.pdf>.

Hartley, D. (2013). Australian NGO Cooperation Program: Meta-Evaluation. <http://dfat.gov.au/about-us/publications/Documents/ancp-meta-evaluation-2013.pdf>.

ICF Macro, Inc. (2012). Evaluation of GAVI Gender Policy Final Report. <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.392.6755&rep=rep1&type=pdf>.

IMF. (2013a). Implementation plan in response to board-endorsed recommendations for the IEO evaluation of international reserved – IMF concerns and country perspectives. <http://www.imf.org/external/np/pp/eng/2013/032113.pdf>.

IMF. (2013b). Implementation plan in response to board-endorsed recommendations for the IEO evaluation of the role of the fund as trusted advisor. <http://www.imf.org/external/np/pp/eng/2013/052913.pdf>.

IMF. (2014a). Sixth periodic monitoring report on the status of implementation plans in response to board-endorsed IEO recommendations. <http://www.imf.org/external/np/pp/eng/2014/063014a.pdf>.

IMF, World Bank. (2006). GLOBAL MONITORING REPORT 2006. <http://www.imf.org/external/pubs/ft/gmr/2006/eng/gmr.pdf>.

Independent Evaluation Group. (2011). Results and Performance of the World Bank Group. <http://www.oecd.org/derec/worldbankgroup/49450796.pdf>.

Independent Evaluation Group. (2012a). Comparison of the M&E Systems of the World Bank and the Global Fund. <http://ieg.worldbankgroup.org/Data/reports/gfr_wp_april2012.pdf>.

Independent Evaluation Group. (2012b). World Bank Group Impact Evaluations: Relevance and Effectiveness. <http://ieg.worldbankgroup.org/Data/reports/impact_eval_report.pdf>.

Independent Evaluation Group. (2013a). Assessing the Monitoring and Evaluation Systems of IFC and MIGA: Biennial Report on Operations Evaluation <http://ieg.worldbank.org/Data/reports/broe_eval_0.pdf>.

Independent Evaluation Group. (2013b). Results and Performance of the World Bank Group 2013: An Independent Evaluation - Volume II: Appendixes. <https://ieg.worldbankgroup.org/Data/reports/chapters/rap2013_vol2updated.pdf>.

Independent Evaluation Office of the IMF. (2011). Research at the IMF: Relevance and Utilization. <http://www.ieo-imf.org/ieo/files/completedevaluations/1.1%20Main%20Report.pdf>.

Independent Evaluation Office of the IMF. (2013). Annual Report 2013. <http://www.oecd.org/derec/imf/IMF_AnnualReport2013.pdf>.

Independent Evaluation Office of the IMF. (2014a). Annual Report 2014. <http://www.ieo-imf.org/ieo/files/annualreports/AR_2014_IEO_full%20report.pdf>.

Independent Evaluation Office of the IMF. (2014b). IMF Forecasts: Process, Quality, and Country Perspectives. <http://www.ieo-imf.org/ieo/files/completedevaluations/Full%20Report.pdf>.

Independent Evaluation Office of the IMF. (2014c). Recurring Issues from a Decade of Evaluation: Lessons for the IMF. <http://www.ieo-imf.org/ieo/files/completedevaluations/full%20text.pdf>.

Independent Evaluation Office of the IMF. (2014d). Revisiting the IEO Evaluations of The IMF's Role in PRSPs and the PRGF (2004) and the IMF and Aid to Sub-Saharan Africa (2007). <http://www.ieo-imf.org/ieo/files/updates/2004%20PRSPs-PRGFs%20and%202007%20Aid%20to%20SSA.pdf>.

Independent Evaluation Office of the IMF. (2015b). PROGRESS REPORT TO THE INTERNATIONAL MONETARY AND FINANCIAL COMMITTEE ON THE ACTIVITIES OF THE INDEPENDENT EVALUATION OFFICE OF THE IMF. <http://www.ieo-imf.org/ieo/files/reportstotheimfc/IMFC_2015__Progress_Report_to_the_IMFC_on_IEO_Activities.PDF>.

Itad. (2014b). Can We Demonstrate the Difference that Norwegian Aid Makes? Evaluation of results measurement and how this can be improved: Annexes 4-15. <http://www.oecd.org/derec/norway/Annex%204-6-Can-we-demonstrate-the-difference-that-Norwegian-aid-makes.pdf>.

JICA. (2013). Annual Report on Japan ODA Evaluation 2013. <http://www.oecd.org/derec/japan/Annual%20Report%20on%20Japan%20ODA%20Evaluation%202013.pdf>.

Jones, R. & Young, V. (2004). CIDA Evaluation Guide. <http://www.enterprise development.org/page/download?id=2100>.

Kluyskens, J. & Faubert, C. (2013). Independent Assessment of the United Nations Evaluation Group 2004-2012. <http://www.uneval.org/document/download/1816>.

Macro International Inc. (2008). Evaluation of the Global Fund Partner Environment, at Global and Country Levels, in Relation to Grant Performance and Health Systems Effects, Including 16 Country Studies. (Contract #GVA-520-C-07-045000) Retrieved from <theglobalfund.org>.

Ministry of Foreign Affairs of Denmark. (2012a). Annual Report 2012 Evaluation of Development Assistance. <http://um.dk/en/~/media/UM/Danish-site/Documents/Danida/Resultater/Eval/EVALredgor2012eng.pdf>.

Ministry of Foreign Affairs of Denmark. (2013). Evaluation programme for 2013-2014). <http://um.dk/en/~/media/UM/Danish-site/Documents/Danida/Resultater/Eval/Programme20132014.pdf>.

Ministry of Foreign Affairs of Denmark. (2015). Eval News. <http://um.dk/~/media/UM/Danish-site/Documents/Danida/Resultater/Eval/EvalNews201501.pdf>.

Ministry of Foreign Affairs of Japan. (2013). ODA Evaluation Guidelines. <http://www.mofa.go.jp/policy/oda/evaluation/basic_documents/pdfs/guidelines2013.pdf>.

Ministry of Foreign Affairs of the Netherlands. (n.d.). Results in Development 2009-2010. <http://www.dsw.org/fileadmin/content/Docs/NL/results-in-development-2009-2010.pdf>.

Multilateral Organization Performance Assessment Network. (2013). Institutional Report: World Food Program. <http://www.mopanonline.org/publications/MOPAN_2013-_WFP_Vol._I.pdf>.

Naudet, J. & Delarue, J. (2007). Fostering impact evaluations at the Agence Française de Développement. <http://www.afd.fr/webdav/site/afd/shared/PUBLICATIONS/RECHERCHE/Evaluations/Notes-methodologiques/02-VA-notes-methodologiques.pdf>.

Norad Evaluation Department. (2014). Evaluation Programme 2014-2016. <http://www.norad.no/globalassets/import-2162015-80434-am/www.norad.no-ny/filarkiv/vedlegg-til-publikasjoner/evaluation-programme-2014-2016.pdf>.

Norad. (2012). Annual Report 2012: Evaluation of Norwegian Development Cooperation. <http://www.oecd.org/derec/norway/Annual-Report-2012.pdf>.

Norad. (2014). Implementation Schedule for the Busan Common Standard. <www.norad.no/en/tools-and-publications/aid-transparency/_attachment/414581%3F_ts%3D1476264566a%26download%3Dtrue+&cd=47&hl=en&ct=clnk&gl=us>.

Ocampo, J.A., Pickford, S., & Rustomjee, C. (2013). External Evaluation of the Independent Evaluation Office. <http://www.imf.org/external/np/pp/eng/2013/010113.pdf>.

OECD & UNDP. (2014). Making Development Cooperation More Effective 2014. <http://effectivecooperation.org/progress>.

OECD. (2010). Evaluation in Development Agencies. <http://www.oecd.org/dac/evaluation/dcdndep/48888783.pdf>.

OECD. (2010). Germany: Development Assistance Committee (DAC) Peer Review. <http://www.oecd.org/dac/peer-reviews/46439355.pdf>.

OECD. (2011a). Aid Effectiveness 2011: Progress in Implementing the Paris Declaration. *OECD Publishing*.

OECD. (2011b). The Netherlands Development Assistance Committee (DAC) Peer Review 2011. <http://www.oecd.org/dac/peerreviews/netherlands>.

OECD. (2012). Canada: Development Assistance Committee (DAC) Peer Review 2012. <http://www.oecd.org/dac/peer-reviews/canadapeerreview2012.pdf>.

OECD. (2013b). Peer Review: Australia 2013. <http://www.oecd.org/dac/peer-reviews/OECD%20Australia%20FinalONLINE.pdf>.

Office of Evaluation and Oversight, IDB. (2014a). Annual Report 2013-2014 Office of Evaluation and Oversight. <http://www.iadb.org/en/office-of-evaluation-and-oversight/ove-annual-report-2013-2014,18716.html>.

Peter Hansen, Abdallah Bchir, Laura Stormont. (2012). Report to the GAVI Alliance Board 12-13 June 2012. G.A.V.I. retreived on April 22, 2015 from <http://www.gavi.org/About/Governance/GAVI-Board/Minutes/2012/12-june/Minutes/01h---Evaluation-policy/>.

Poate, D. & Barnett, C. (2003). Measuring Value for Money? An Independent Review of DFID's Value for Money (VFM) Indicator, Public Service Agreement 2003-2006. <https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/67941/measuringvalueformoney>.

Shorten, T., et al. (2012). The International Health Partnership Plus: rhetoric or real change? Results of a self-reported survey in the context of the 4th high level forum on aid effectiveness in Busan. *Globalization and Health*, 8(13).

Spitz, Gabi, Muskens, Roeland, & vn Ewijk, Edith, NCDO. (2013). The Dutch and development cooperation: Ahead of the crowd or trailing behind? Amsterdam.

Swedish Agency for Development Evaluation. (2010). Evaluation of Sweden's Implementation of the 2010 Paris Declaration. <http://www.oecd.org/derec/sweden/47681398.pdf>.

Swedish International Development Cooperation Agency. (2007). Looking Back, Moving Forward: Sida Evaluation Manual. <http://www.enterprise-development.org/page/download?id=2026>.

Technical Evaluation Reference Group. (2006). Summary Report of the Technical Evaluation Reference Group (TERG) of the Global Fund; Fourth Meeting. Glion, Switzerland 27 Feb-1 Mar 2006 Retrieved from <theglobalfund.org>

Technical Evaluation Reference Group. (2009). Study Area 3 HEALTH IMPACT OF SCALING UP AGAINST HIV, Tuberculosis AND MALARIA: EVALUATION OF THE CURR ENT SITUATION AND TRENDS IN 18 COUNTRIES. The Global Fund. Retrieved from <theglobalfund.org>

Technical Evaluation Reference Group. (2009). Study Area 3 HEALTH IMPACT OF SCALING UP AGAINST HIV, Tuberculosis AND MALARIA: EVALUATION OF THE CURR ENT SITUATION AND TRENDS IN 18 COUNTRIES. The Global Fund. Retrieved from <theglobalfund.org>.

The Global Fund. (2009). Affordable medicines Facility - Malaria (phase 1) monitoring & evaluation technical framework. Retrieved from <theglobalfund.org>.

The Global Fund. (2011). Key Performance Indicators: End Year Results for 2011; 2nd Strategy, Investment and Impact & Finance and Operational Performance Committee Meetings. Geneva, 27-28 March 2012. Retrieved from <theglobalfund.org>.

The Global Fund. (2012). Key Performance Indicators: End Year Results for 2012. Retrieved from <theglobalfund.org>.

The Global Fund. (2012). The Global Fund Annual Report. <http://www.theglobalfund.org/en/publications/2013-07-24_Annual_Report_2012/>.

The Global Fund. (2012). The Global Fund Strategy 2012-2016: Investing for Impact. Retrieved from <theglobalfund.org>

The Global Fund. (2013). The Global Fund Strategy Corporate key Performance Indicator Framework for 2014-2016. Retrieved from <theglobalfund.org>.

UK National Audit Office. (2012). The multilateral aid review. <http://www.nao.org.uk/wp-content/uploads/2012/09/1213594.pdf>.

United Nations Evaluation Group. (2005). Norms for Evaluation in the UN System.
        <http://www.uneval.org/document/detail/21>.

United Nations Evaluation Group. (2013a). UNEG Strategy 2014–2019. <http://www.uneval.org/document/detail/1459>.

United Nations Evaluation Group. (2013b). Impact Evaluation in UN Agency Evaluation Systems: Guidance on Selection,
        Planning, and Management. <http://www.uneval.org/document/download/1880>.

United States Department of State. (2014). 2014 Standard Foreign Assistance Indicators.
        <http://www.state.gov/f/indicators/>.

USAID. (2011). Evaluation: Learning from Experience: USAID Evaluation Policy.
        <http://www.usaid.gov/sites/default/files/documents/2151/USAIDEvaluationPolicy.pdf>.

USAID. (2012). ADS Chapter 203: Assessing and Learning.
        <http://www.usaid.gov/sites/default/files/documents/1870/203.pdf>.

USAID. (2013a). Annual Performance Monitoring Report FY13.
        <https://www.usaidassist.org/sites/assist/files/usaid_assist_project_annual_performance_monitoring_report_fy_1
        3.pdf>.

USAID. (2013b). Second Food Aid and Food Security Assessment (FAFSA-2): Performance Management.
        <http://www.fantaproject.org/sites/default/files/resources/FAFSA-2-Chapter9-Mar2013.pdf>.

USAID. (2013g). FOREIGN OPERATIONS FY 2013 Performance Report. <http://www.usaid.gov/results-and-data/progress-
        data/annual-performance-report>.

USAID. (2013h). Performance Management Plan (PMP) Toolkit. <http://pdf.usaid.gov/pdf_docs/pa00jhcb.pdf>.

USAID. (2014). AGENCY FINANCIAL REPORT FISCAL YEAR 2014. <http://www.usaid.gov/results-and-data/progress-
        data/agency-financial-report>.

Van Kerkhoff, L., & Szlezák, N. (2006). Linking local knowledge with global action: examining the Global Fund to Fight AIDS,
        Tuberculosis and Malaria through a knowledge system lens. *Bulletin of the World Health Organization*, 84(8), 629-
        635.

World Bank. (2014). World Bank Annual Report. <www.worldbank.org/en/about/annual-report>.

World Food Program. (2008). WFP Evaluation Policy. <http://one.wfp.org/eb/docs/2008/wfp187763~2.pdf>.

World Food Program. (2013). Annual Evaluation Report.
        <http://documents.wfp.org/stellent/groups/public/documents/reports/wfp264960.pdf>.

World Health Organization (WHO). (n.d). Country monitoring and evaluation guidance. Retrieved 14 May 2015 from
        <http://www.who.int/healthinfo/country_monitoring_evaluation/en/1/>.

World Health Organization (WHO). (2014b). Twelfth General Programme of Work 2014–2019
<http://apps.who.int/iris/bitstream/10665/112792/1/GPW_2014-2019_eng.pdf?ua=1>.

Appendix A. Literature Search Process

Our search aimed to identify documents describing or reviewing the performance of major international aid donor organizations and these organizations' systems for measuring performance. We conducted a Google search to identify relevant non-published literature and key multilateral organizations to review, and also conducted searches of PAIS International and Scopus to identify relevant published literature. In addition, we searched the websites of 12 bilateral and ten multilateral donor organizations. All of these searches used the following search string: *(system OR framework) AND ("economic development" OR "international development" OR aid) AND ("monitoring and evaluation" OR "performance management" OR "performance monitoring" OR "results management" OR "results data" OR "value for money")*. This search string was designed to limit search results to results measurement systems relevant to international aid and to capture multiple types of documents referring to systems designed to monitor and evaluate results.

For all of our searches, we screened the titles and summaries of the search results and retrieved all documents that met the following screening criteria:
- Reports on the results or performance measurement of at least one specific donor organization;
- Reports on organizational performance OR the system and processes for measuring performance;
- Reports on measurement of aggregate organizational-level performance (as opposed to project-, recipient-country-, or portfolio-level performance); and
- Availability in full-text in English.

The evidence we gathered does include information on donor results measurement processes at the project, country, and portfolio levels, but we excluded documents that focused only on these sub-levels to narrow the scope of the review and to emphasize organizational-level results measurement.

Our Google search yielded over 30 million results, though these were decreasingly relevant after the first 100. We reviewed the first 200 results, and retrieved 29 relevant documents, including annual reports, reviews, and evaluations of major bilateral and multilateral donors. We also followed links to the Donor Committee for Enterprise Development and OECD DAC Evaluation Resource Centre websites and retrieved an additional 24 relevant documents. Our searches of academic databases indicated that there is little published information examining donor organizations' results measurement. We reviewed all 227 search results on PAIS International and all 301 results on Scopus but only retrieved two relevant documents from PAIS and one from Scopus.

We performed targeted searches for the ten largest bilateral donors in terms of 2014 ODA (OECD, 2014d), as well as two other bilateral donors that appeared in our initial Google search (Austria, Denmark). We reviewed the results measurement systems of the main aid organizations of each of these 12 bilateral donors. In addition, we performed targeted searches for seven of the ten largest multilateral donor organizations in terms of 2013 multilateral aid flows (OECD, 2014c). We did not include the European Bank for Reconstruction and Development (EBRD), and instead reviewed the three larger development banks, the ADB, AfDB, and IADB. We also did not include EU Institutions or Arab Funds, as these are groupings of several smaller organizations which are difficult to review as a unit. In addition, we also review the GAVI Alliance, the 11th largest multilateral donor, as well as two agencies of the United Nations Group[28].

For each of these targeted donor organizations, we searched their websites with the same general search string as above. In addition to these searches, we browsed the websites' publications, documents, reports, results, or evaluation pages to identify any relevant recent documents that did not come up in the search results. These searches targeted reports that demonstrated the types of results that donors measure and report, as well as monitoring or evaluation plans describing the plans and processes for measuring results. We also conducted targeted Google searches for peer reviews of all donor organizations.

These targeted searches yielded the following numbers of additional relevant documents:
- United States Agency for International Development (USAID), 10 documents
- United Kingdom Department for International Development (DFID), 9 documents
- Deutsche Gesellschaft für Internationale Zusammenarbeit (GIZ), 10 documents
- Agence Française de Développement (AFD), 13 documents

---

[28] The World Food Programme (WFP) and World Health Organization (WHO) are United Nations agencies. Our coding of United Nations Group is based on system-wide characteristics, but we considered evidence on the WHO and WFP separately, in order to assess whether there are differences in results measurement for the more specialized agencies.

- Japan International Cooperation Agency (JICA), 5 documents
- Swedish International Development Cooperation Agency (Sida), 3 documents
- Dutch Ministry of Foreign Affairs (MFA), 3 documents
- Norwegian Agency for Development Cooperation (Norad), 4 documents
- Australia Department of Foreign Affairs and Trade (DFAT), 10 documents
- Canada Department of Foreign Affairs, Trade and Development (DFATD), 12 documents
- Austria Development Cooperation (ADC), 4 documents
- Ministry of Foreign Affairs of Denmark (Danida), 8 documents
- African Development Bank (AfDB), 8 documents
- Asian Development Bank (ADB), 9 documents
- GAVI Alliance (GAVI), 14 documents
- Global Fund, 14 documents
- Inter-American Development Bank (IADB), 4 documents
- International Monetary Fund (IMF), 14 documents
- United Nations Group (UNG), 10 documents
- World Health Organization (WHO), 4 documents
- World Bank Group (WBG), 5 documents
- World Food Programme (WFP), 5 documents

We retrieved a total of 234 relevant documents from these searches[29]. These documents primarily consist of donors' annual reports, evaluation plans, policies for monitoring, evaluation, and performance measurement, and internal evaluations of effectiveness and performance. We also retrieved peer reviews of many donors, mostly facilitated by the Organization for Economic Cooperation and Development (OECD)'s Development Assistance Committee (DAC). In addition, we retrieved several documents reporting on the performance of multiple donor organizations. The multi-country documents include:
- An OECD report on making development cooperation more effective (OECD & UNDP, 2014);
- An OECD report on measuring and managing results in development cooperation (OECD, 2014b);
- An OECD overview of development results measurement and management (OECD, 2013c);
- An OECD report on progress in implementing the Paris Declaration for aid effectiveness (OECD, 2011);
- A review of multilateral aid organizations (UK National Audit Office, 2012) and update to the review (DFID, 2013);
- An article on health aid effectiveness (Shorten, et al., 2012);
- An article on development evaluation doctrine and how evaluate aid (Armytage, 2011);
- A review of the performance and governance of multilateral donor organizations (Vivideconomics, 2011);
- An article on developing an aid effectiveness index (Sinha, 2010);
- A 2000 OECD DAC review of results-based management in the development co-operation agencies (Binnendijk, 2000).

Figure A.1 displays the proportions of bilateral ODA provided by members of the Organization for Economic Cooperation and Development's (OECD) Development Assistance Committee (DAC) in 2013 (OECD, 2014c). The 12 donor countries we have selected for review represent 87 percent of total DAC ODA in 2013. *Figure A.2* presents the proportion of multilateral aid flows[30] provided by the ten multilateral organizations we selected. These ten organizations represent 75 percent of multilateral aid (*ibid.*).

---

[29] Of the original 254 documents we retrieved, 13 were duplicates and 19 proved not to be relevant after further review.
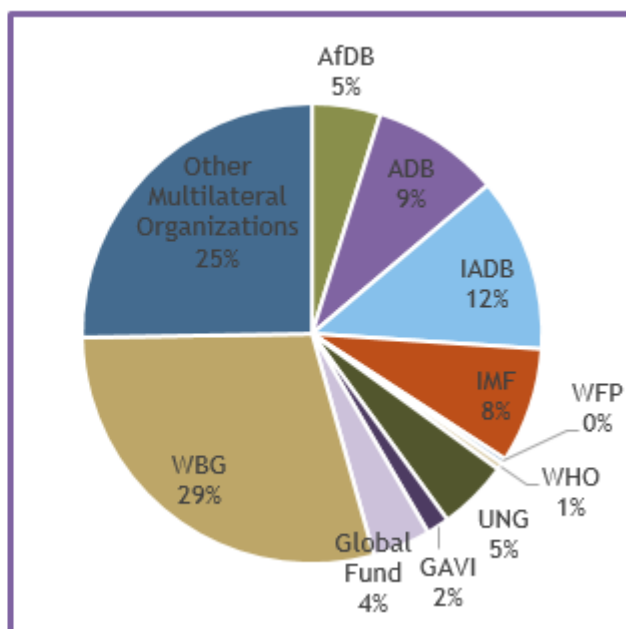[30] Multilateral aid flows are calculated as the sum of concessional and non-concessional by multilateral organizations is US$ millions.

Figure A.1: Proportion of Total Bilateral ODA from Selected OECD DAC Member Countries, 2013
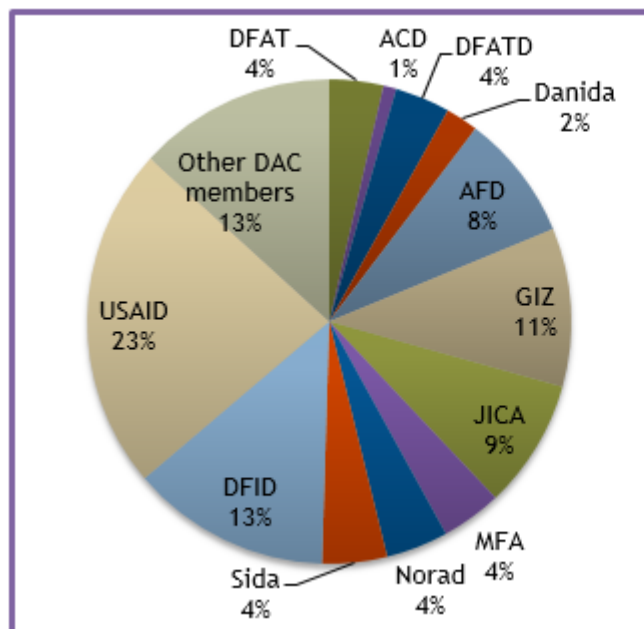
Figure A.2: Proportion of Total MultilateralAid Flows from Selected Multilateral Organizations, 2013

Source: OECD, 2014c.

Appendix B. Summary of Body of Evidence

Table B.1 summarizes the body of evidence for each of the 22 donor results measurement systems we reviewed. The table outlines the total number of documents reviewed as well as the number of external documents reviewed, in order to illustrate whether the evidence comes from only the organizations themselves or is supported by external reviews. The table also presents the extent to which the evidence we reviewed is recent. For each donor organization, we list the year of the most recent annual report or review of performance and results, the year of the most recent review of their evaluation or performance measurement system, and the year of the most recent peer review, if available.
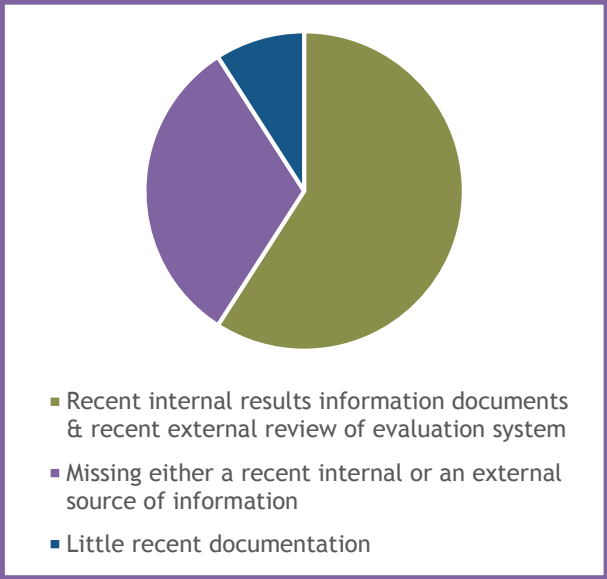
We also rate the strength of the body of evidence for each results measurement system based on the relevance and quality of the information included in the documents. A "high" rating (13 of 22 donors) indicates that both a recent internal document reporting results information (e.g. an annual report/review) and a recent external review of the organization's evaluation system is available. Bodies of evidence we rate "medium" (seven of 22 donors) are missing either a recent internal or an external source of information. We rate two bodies of evidence for which little recent documentation is available "low" (Dutch MFA, ADC). *Figure B.1* illustrates the proportion of the bodies of evidence for donor results measurement systems rated "high," "medium," and "low".

Table B.1 Summary of Body of Evidence for Each Government M&E System

Donor Organization	Bilateral or Multilateral	Total Number of Documents Reviewed	Number of External Documents Reviewed	Year of Most Recent Annual Report/Review	Year of Most Recent Review of Evaluation/ Performance Measurement System	Year of Most Recent Peer Review	Strength of Body of Evidence
AFD	Bilateral	12	3	2013	2013	2013	High
AfDB	Multilateral	12	4	2014	2013	NA	High
ADB	Multilateral	14	4	2015	2014	NA	High
ADC	Bilateral	5	3	NA	2009	2009	Low
Danida	Bilateral	8	3	NA	2014	2014	Medium
DFAT	Bilateral	18	5	2015	2013	2013	High
DFATD	Bilateral	19	4	2014	2012	2012	High
DFID	Bilateral	10	5	2014	2014	2014	High
MFA	Bilateral	5	4	NA	2011	2011	Low
GAVI	Multilateral	12	5	2013	2011	NA	Medium
GIZ	Bilateral	11	3	2012	2012	NA	High
Global Fund	Multilateral	18	5	2012	2011	NA	Medium
IADB	Multilateral	9	5	2014	2013	NA	High
IMF	Multilateral	16	0	2014	2014	NA	Medium
JICA	Bilateral	8	3	2014	2014	NA	High
Norad	Bilateral	12	6	2013	2014	NA	High
Sida	Bilateral	8	6	NA	2013	2013	Medium
UNG	Multilateral	10	5	NA	2013	NA	Medium
USAID	Bilateral	16	7	2014	2013	NA	High
WHO	Multilateral	6	2	NA	2011	NA	Medium
WBG	Multilateral	12	5	2013	2014	NA	High
WFP	Multilateral	6	3	2013	2014	2014	High

Figure B.1: Strength of Bodies of Evidence

- Recent internal results information documents & recent external review of evaluation system
- Missing either a recent internal or an external source of information
- Little recent documentation

Appendix C. Review Framework Questions

Basic System Descriptives
- Name of organization
- Acronyms used
- Number of documents reviewed
- Number of internal (organizational) documents reviewed
- List of internal documents (title and year)
- Number of external (non-organizational) documents reviewed
- List of external documents (title and year)
- Documents include an organizational annual report or review?
- Year of most recent organizational annual report or review
- Documents include a report of review of the organization's evaluation or results measurement system?
- Year of most recent report or review of the organization's evaluation or results measurement system
- Documents include a peer review of the organization?
- Year of most recent peer review

Institutional Design
- What office is primarily responsible for evaluating performance?
- Is there a periodic evaluation plan?
- Are evaluations conducted by internal evaluators?
- Is the internal evaluation group independent of implementing agencies?
- Are evaluations conducted by independent external evaluators?
- Are there challenges with or recommendations for the capacity for evaluation (e.g. analytical skills, staffing, knowledge of theory of change, communication with evaluation stakeholders)?
- Evaluation Guidelines
    - Does the donor organization require a theory of change, results chain, or results framework for its programs/grants?
    - Does the donor organization provide guidelines or standards for data collection or aggregation for its programs/grants?
    - Does the donor organization provide guidelines or standards for data verification or data quality for its programs/grants?
    - Who performs data verification or oversees data quality?
    - Does the donor organization provide guidelines or standards for reporting for its programs/grants?

Level of Evaluation
- Is there a process for reviewing and evaluating grant/project-level performance?
    - Are there challenges with or recommendations for collecting or reviewing grant/project-level results information (e.g. lack of baseline or other data)?
- Is there a process for reviewing and evaluating grantee/partner-level performance?
    - Are there challenges with or recommendations for collecting or reviewing grantee/partner-level results information (e.g. lack of baseline or other data)?
- Is there a process for reviewing and evaluating recipient-country-level performance?
    - Are there challenges with or recommendations for collecting or reviewing recipient-country-level results information (e.g. different indicators, data aggregation issues, lack of baseline data)?
- Is there a process for reviewing and evaluating portfolio-level performance (e.g. sectoral or regional)?
    - Are there challenges with or recommendations for collecting or reviewing portfolio-level results information (e.g. different indicators, data aggregation issues, lack of baseline data)?
- Is there a process for reviewing and evaluating aggregate performance (e.g. across all grants)?
    - Are there challenges with or recommendations for collecting or reviewing aggregate organizational results information (e.g. different indicators, data aggregation issues, lack of baseline data)?

Results: Coordination and Alignment with Recipients
- Are there challenges with or recommendations for coordination with grantees/implementing partners (e.g. aligning indicators, data collection, reporting)?

- What is the overall level of harmonization of donor and recipient government aid systems (overall rating derived from below indicators)?
- How does the organization rate overall according to the targets for aid effectiveness in the 2011 OECD review of aid effectiveness? (low to high)
    - How much technical assistance is coordinated with recipient-country programs?
    - How much aid uses recipient-country systems?
    - How much bilateral aid is untied?
    - How many PIUs are parallel to recipient-country structures?
    - How many donor missions were coordinated?
    - How many recipient-country analyses were coordinated?
- How does the organization rate overall according to the indicators in Sinha's (2010) aid effectiveness review? (low to high)
    - What is the organization's rating for the Strengthen Capacity by Coordinated Support (SCCS) Index?
    - What is the organization's rating for the Use of Country System Index (Public Financial Management – PFM, Procurement)?
    - What is the organization's rating for the Avoidance of Parallel Implementation Structures Index?
    - What is the organization's rating for the Predictability of Aid Index?
    - What is the organization's rating for the Untying of Aid Index?
    - What is the organization's rating for the Common Arrangement Index?
    - What is the organization's rating for the Shared Analysis Index?
    - What is the organization's rating for the Overall Aid Effectiveness Index?
    - How does the organization rank among the organizations reviewed by Sinha (2010)?
    - What are the relative strengths and weaknesses of the organization according to these indicators?

Results: Outputs and Implementation
- Do evaluations assess performance in terms of activities completed or outputs achieved?
- Do evaluations compare output results against targets to assess performance?
- Do evaluations compare output results across projects to assess performance?
- Does the donor organization specify the collection of any particular output indicators?
- Does the donor organization specify the collection of outputs disaggregated by gender?
- Are data on specified indicators often actually collected?
- Are data on specified gender-disaggregated output indicators often actually collected?
- Are there challenges with or recommendations for collecting information about outputs of donor-funded projects?

Results: Outcomes and Impact
- Do evaluations assess performance in terms of outcomes resulting from donor-funded projects?
- Do outcome/impact evaluations make connections with projects' theory of change or results framework?
- Do evaluations compare outcome results against targets to assess performance?
- Do evaluations compare outcome results against baseline measures to assess performance?
- Do evaluations compare outcome results against control groups to assess performance?
- Do evaluations compare outcome results against control groups to assess performance often?
- Do evaluations compare outcome results across projects to assess performance?
- Does the donor organization specify the collection of any particular outcome indicators?
- Does the donor organization specify the collection of outcomes belonging to specific categories (e.g. gender, climate, MDGs)?
- Are there challenges with or recommendations for collecting information about outcomes of donor-funded projects?

Results: Costs and Effectiveness
- Do evaluations assess performance in terms of costs of donor-funded projects?
- Do evaluations compare costs against outputs to assess implementation performance?
- Do evaluations compare costs/expenditures against budgets to assess implementation performance?
- Do evaluations compare costs against changes in outcomes to assess cost-effectiveness?
- Do evaluations compare cost-effectiveness results across projects to assess performance?
- Are there challenges with or recommendations for collecting information about costs of donor-funded projects?

Reporting and Use of Results Information
- Are there documented rules for reporting donor results information (at any level)?
- Are there documented rules for reporting organization-level results?
- Is results information shared within the donor organization?
- Is results information shared with grantees/implementing partners?
- Is results information disseminated to the general public?
- Are there challenges with or recommendations for sharing results information?
- Is there an an established process or incentive system for incorporating results information?
- Does results information influence strategy, planning, or project design or operations?
- Does results information influence budget allocations (e.g. results-based funding)?
- Who is responsible for tracking whether the organization acts on M&E information?
- Are there processes for recording failure in projects (e.g. negative impacts or smaller impacts than anticipated) and incorporating lessons about failure?
- Are there challenges with or recommendations for incorporating results information?
- Are there other challenges with or recommendations for the use of results information?

Appendix D. OECD DAC Quality Standards for Development Evaluation[31]

	Standard	Definition	Representation in Review Framework (including specific review framework questions)	Relevant Section of Report
1. Overarching Considerations	1.1 Free and open evaluation process	The evaluation process is transparent and independent from programme management and policy-making, to enhance credibility.	• Is the internal evaluation group independent of implementing agencies? • Are evaluations conducted by independent external evaluators?	4.1.1
	1.2 Evaluation ethics	Evaluation abides by relevant professional and ethical guidelines and codes of conduct for individual evaluators.	Not represented	N/A
	1.3 Partnership approach	The concept of partnership connotes an inclusive process, involving different stakeholders such as government, parliament, civil society, intended beneficiaries and international partners.	• What is the overall level of harmonization of donor and recipient government aid systems? • Are there challenges with or recommendations for coordination with grantees/implementing partners (e.g. aligning indicators, data collection, reporting)?	4.2, 4.3
	1.4 Co-ordination and alignment	The evaluation process takes into account national and local evaluation plans, activities and policies.	• How many donor missions were coordinated? • How many recipient-country analyses were coordinated?	4.3.1
	1.5 Capacity development	Positive effects of the evaluation process on the evaluation capacity of development partners are maximized.	• Are there challenges with or recommendations for the capacity for evaluation?	4.1.2, 4.1.4
	1.6 Quality control	Depending on the evaluation's scope and complexity, quality control is carried out through an internal and/or external mechanism.	• Does the donor organization provide guidelines or standards for data verification or data quality for its programs/grants? • Who performs data verification or oversees data quality? • Does the donor organization provide guidelines or standards for reporting for its programs/grants?	4.1.3
2. Purpose, planning and design	2.1 Evaluation object and scope	The development intervention being evaluated (the evaluation object) is clearly defined, including a description of the intervention logic or theory. The evaluation scope defines the time period, funds spent, geographical area, target groups, organisational set-up, implementation arrangements, policy and institutional context and other dimensions to be covered by the evaluation.	• Is there a periodic evaluation plan? • Does the donor organization require a theory of change, results chain, or results framework for its programs/grants? • Do outcome/impact evaluations make connections with projects' theory of change or results framework?	4.1.1, 4.1.3, 4.5
	2.2 Evaluability	The feasibility of an evaluation is assessed. Specifically, it should be determined whether or not the development intervention is adequately defined and its results verifiable, and if evaluation is the	Not represented – refers to specific evaluation projects rather than overall results measurement processes	N/A

---

[31] All standards and definitions are taken from the OECD DAC's 2010 publication "Quality Standards for Development Evaluation."

		best way to answer questions posed by policy makers or stakeholders.		
	2.3 Stakeholder involvement	Relevant stakeholders are involved early on in the evaluation process and given the opportunity to contribute to evaluation design, including by identifying issues to be addressed and evaluation questions to be answered.	Not represented – the review framework considers evaluation guidelines from the perspective of the donor organization, but does not capture information on stakeholder involvement in preparing evaluations.	N/A
	2.4 Systematic consideration of joint evaluation	To contribute to harmonisation, alignment and an efficient division of labour, donor agencies and partner countries systematically consider the option of a joint evaluation, conducted collaboratively by more than one agency and/or partner country.	• How many donor missions were coordinated? • How many recipient-country analyses were coordinated?	4.3.1
	2.5 Evaluation questions	The evaluation objectives are translated into relevant and specific evaluation questions.	Not represented – refers to specific evaluation projects rather than overall results measurement processes	N/A
	2.6 Selection and application of evaluation criteria	The evaluation applies the agreed DAC criteria for evaluating development assistance: relevance, efficiency, effectiveness, impact and sustainability.	Our review framework divides results into four categories. Efficiency is represented by questions about comparing costs against outputs and outcomes. Effectiveness, impact, and sustainability are represented by questions about measuring performance in terms of outcome results. Relevance is not represented.	4.5, 4.6
	2.7 Selection of approach and methodology	The purpose, scope and evaluation questions determine the most appropriate approach and methodology for each evaluation.	Not represented – refers to specific evaluation projects rather than overall results measurement processes	N/A
	2.8 Resources	The resources provided for the evaluation are adequate, in terms of funds, staff and skills, to ensure that the objectives of the evaluation can be fulfilled effectively.	• Are there challenges with or recommendations for the capacity for evaluation? • Are there challenges with or recommendations for collecting or reviewing results information at different levels?	4.1.2, 4.1.4, 4.2
	2.9 Governance and management structures	The evaluation governance structure safeguards credibility, inclusiveness, and transparency. Management organises the evaluation process and is responsible for day-to-day administration.	• What office is primarily responsible for evaluating performance?	4.1.1
	2.10 Document defining purpose and expectations	The planning and design phase culminates in the drafting of a clear and complete written document, usually called "Terms of Reference" (TOR).	• Does the donor organization require a theory of change, results chain, or results framework for its programs/grants? • Does the donor organization provide guidelines or standards for data collection or aggregation for its programs/grants? • Does the donor organization provide guidelines or standards for data verification or data quality for its programs/grants?	4.1.3

3. Implementation and reporting	3.1 Evaluation team	A transparent and open procurement procedure is used for selecting the evaluation team. The members of the evaluation team possess a mix of evaluative skills and thematic knowledge.	• Are evaluations conducted by internal evaluators? • Is the internal evaluation group independent of implementing agencies? • Are evaluations conducted by independent external evaluators? • Are there challenges with or recommendations for the capacity for evaluation?	4.1
	3.2 Independence of evaluators vis-à-vis stakeholders	Evaluators are independent from the development intervention, including its policy, operations and management functions, as well as intended beneficiaries.	• Is the internal evaluation group independent of implementing agencies? • Are evaluations conducted by independent external evaluators?	4.1.1
	3.3 Consultation and protection of stakeholders	The full range of stakeholders, including both partners and donors, are consulted during the evaluation process and given the opportunity to contribute.	Not represented – the review framework considers results measurement from the perspective of the donor organization, but does not systematically capture information on stakeholder involvement in conducting evaluations.	N/A
	3.4 Implementation of evaluation within allotted time and budget	The evaluation is conducted and results are made available to commissioners in a timely manner to achieve the objectives of the evaluation. The evaluation is carried out efficiently and within budget.	• Are there documented rules for reporting organization-level results? • Is results information shared within the donor organization? • Is results information shared with grantees/implementing partners? • Is results information disseminated to the general public?	4.7.1
	3.5 Evaluation report	The evaluation report can readily be understood by the intended audience(s) and the form of the report is appropriate given the purpose(s) of the evaluation.	• Are there challenges with or recommendations for sharing results information? • Are there challenges with or recommendations for incorporating results information?	4.7
	3.6 Clarity and representative-ness of summary	A written evaluation report contains an executive summary. The summary provides an overview of the report, highlighting the main findings, conclusions, recommendations and any overall lessons.	• Are there documented rules for reporting organization-level results? • Are there challenges with or recommendations for sharing results information? • Are there challenges with or recommendations for incorporating results information?	4.7
	3.7 Context of the development intervention	The evaluation report describes the context of the development intervention.	Our review framework captures how results information is measured and reported at different levels of measurement.	4.2
	3.8 Intervention logic	The evaluation report describes and assesses the intervention logic or theory, including underlying assumptions and factors affecting the success of the intervention.	• Does the donor organization require a theory of change, results chain, or results framework for its programs/grants? • Do outcome/impact evaluations make connections with projects' theory of change or results framework?	4.1.3, 4.5

3.9 Validity and reliability of information sources	The evaluation report describes the sources of information used (documents, respondents, administrative data, literature, etc.) in sufficient detail so that the adequacy of the information can be assessed.	Our review framework captures challenges with different levels and types of results measurement, including challenges with information sources.	Multiple
3.10 Explanation of the methodology used	The evaluation report describes and explains the evaluation methodology and its application.	Not represented	N/A
3.11 Clarity of analysis	The evaluation report presents findings, conclusions, recommendations and lessons separately and with a clear logical distinction between them.	This standard is somewhat represented on a case by case as information was available in the section on reporting and use of results information, through questions about challenges and recommendation for sharing and incorporating results information.	4.7
3.12 Evaluation questions answered	The evaluation report answers all the questions detailed in the TOR for the evaluation.	Not represented	N/A
3.13 Acknowledge-ment of changes and limitations of the evaluation	The evaluation report explains any limitations in process, methodology or data, and discusses validity and reliability. It indicates any obstruction of a free and open evaluation process which may have influenced the findings.	This standard is somewhat represented on a case by case as information was available in the section on reporting and use of results information, through questions about challenges and recommendation for sharing and incorporating results information and about recording failure in projects.	4.7
3.14 Acknowledge-ment of disagreements within the evaluation team	Evaluation team members have the opportunity to dissociate themselves from particular judgements and recommendations on which they disagree.	Not represented	N/A
3.15 Incorporation of stakeholders' comments	Relevant stakeholders are given the opportunity to comment on the draft report.	Not represented	N/A
4. Follow-up, use and learning — 4.1 Timeliness, relevance and use of the evaluation	The evaluation is designed, conducted and reported to meet the needs of the intended users.	• Is there an established processes or incentive system for incorporating results information? • Are there challenges with or recommendations for incorporating results information?	4.7.2, 4.7.3
4.2 Systematic response to and follow-up on recommenda-tions	Recommendations are systematically responded to and action taken by the person(s)/body targeted in each recommendation.	• Is there an established processes or incentive system for incorporating results information? • Does results information influence strategy, planning, or project design or operations? • Does results information influence budget allocations (e.g. results-based funding)?	4.7.2

4.3 Dissemination	The evaluation results are presented in an accessible format and are systematically distributed internally and externally for learning and follow-up actions and to ensure transparency.	• Are there documented rules for reporting organization-level results? • Is results information shared within the donor organization? • Is results information shared with grantees/implementing partners? • Is results information disseminated to the general public? • Are there challenges with or recommendations for sharing results information?	4.7.1

Appendix E. Donor Acronyms

ADB	Asian Development Bank
ADC	Austria Development Cooperation
AFD	Agence Française de Développement (France)
AfDB	African Development Bank
Danida	Ministry of Foreign Affairs of Denmark
DFAT	Australia Department of Foreign Affairs and Trade
DFATD	Canada Department of Foreign Affairs, Trade and Development
DFID	United Kingdom Department for International Development
GIZ	Deutsche Gesellschaft für Internationale Zusammenarbeit (Gemany)
GAVI	GAVI Alliance
IADB	Inter-American Development Bank
IMF	International Monetary Fund
JICA	Japan International Cooperation Agency
MFA	Dutch Ministry of Foreign Affairs (Netherlands)
Norad	Norwegian Agency for Development Cooperation
Sida	Swedish International Development Cooperation Agency
UNG	United Nations Group
USAID	United States Agency for International Development
WBG	World Bank Group
WFP	World Food Programme
WHO	World Health Organization