

What Lies Behind “Good” Analytical Work on Development?

Four Years of Knowledge Products at the World Bank

Martín Rama

Rucheta Singh

Aiza Aslam



WORLD BANK GROUP

Operations Policies and Country Services Vice-Presidency

February 2024

Abstract

The World Bank's analytical work has a strong reputation, but its knowledge products are also perceived to be of varying quality and relevance, and the drivers of this heterogeneity are only partially understood. Building on previous evaluations, this paper adopts a production function approach to assess how budget resources, time to completion, technical skills, and institutional responsibilities affect the internal ratings and external visibility of different types of analytical tasks at the World Bank. To this effect, the paper first matches records from three unconnected electronic platforms—for internal documents, budget codes, and external publications—to assemble a comprehensive database of knowledge products and their key characteristics. With analytical documents as its unit of observation, the exercise shows that: (1) devoting more

resources to analytical tasks leads to both better ratings and greater visibility; (2) both outcomes are systematically worse when a greater share of resources comes from trust funds; (3) they are also consistently worse for tasks that take longer to complete; (4) more academically oriented team leaders underperform on ratings and overperform on visibility, whereas technically solid but less stellar team leaders overperform on ratings; and (5) everything else equal, performance varies systematically with the nature of the unit in charge. The findings of the paper can be read as a cautionary note against knowledge management that is based on the counting of analytical tasks. Instead, the findings call for much stronger information systems on knowledge products, a better alignment of incentives for the units in charge, and regular evaluations in the spirit of this paper.

This paper is a product of the Operations Policies and Country Services Vice-Presidency. It is part of a larger effort by the World Bank to provide open access to its research and make a contribution to development policy discussions around the world. Policy Research Working Papers are also posted on the Web at <http://www.worldbank.org/prwp>. The authors may be contacted at mrma@worldbank.org.

The Policy Research Working Paper Series disseminates the findings of work in progress to encourage the exchange of ideas about development issues. An objective of the series is to get the findings out quickly, even if the presentations are less than fully polished. The papers carry the names of the authors and should be cited accordingly. The findings, interpretations, and conclusions expressed in this paper are entirely those of the authors. They do not necessarily represent the views of the International Bank for Reconstruction and Development/World Bank and its affiliated organizations, or those of the Executive Directors of the World Bank or the governments they represent.

What Lies Behind “Good” Analytical Work on Development? Four Years of Knowledge Products at the World Bank

Martín Rama, Rucheta Singh and Aiza Aslam *

Keywords: Development Economics, Policy Research, Development Effectiveness.

JEL classification: O19, O20, B40, D73.

* The research underlying this paper was initiated in the summer of 2020 at the request of the Operations Policies and Country Services Vice-Presidency of the World Bank (OPSVP). The authors are grateful to Manuela Ferro and Mamta Murthi for guidance, and to Maurizio Bussolo, Stephane Guimbert, Alexander Korolyov and Aart Kraay for useful comments. The views in the paper are those of the authors and do not engage the World Bank. The corresponding author is Martin Rama (mrma@worldbank.org).

1. Introduction

The World Bank prides itself on being a knowledge-based institution, to a much greater extent than most of its peers. And rightly so. The World Bank was one of the first international organizations to appoint a Chief Economist, in 1972; since then, two of the job holders went on to win the Nobel Prize in economics. With the *World Development Report*, a series introduced in 1978, it pioneered the production of regular flagship studies shaping the narrative on development economics. In 1990, it championed an approach to the measurement of poverty – back then, “one dollar a day” – that has become widely used. Its *Doing Business* reports, launched in 2002, prompted lively policy debates on regulation, and triggered uncountable economic reforms around the world. In 2011, it embraced an open access approach, putting a vast majority of its reports and data under a Creative Commons license; today its *World Development Indicators* are among the most widely used cross-country databases worldwide.

In a reputable ranking of top academic institutions in the field of development, the World Bank comes first, followed by its own research department, ahead of the Massachusetts Institute of Technology, the National Bureau of Economic Research, and other prestigious economics departments (RePEc 2022). And operational counterparts in developing countries often report that they value the knowledge they get from the World Bank more than the financial resources it provides (Knack et al. 2020).

Few other international organizations – if any – may match this record. And yet, even seasoned insiders would be hard pressed to explain how the World Bank accomplishes all this. How is rigorous analytical work produced, what ensures that it is relevant to development practitioners, and how it reaches its intended audience seem to be more art than science.

This is not for lack of trying. The phrase “the knowledge Bank” was coined in 1996, and it is certainly catchy. But beyond obvious themes, such as tapping the World Bank’s global expertise, blending development knowledge with policy engagements, and leveraging digital technologies, it is thin on the specifics (Forbes 2020). Over the half-century since the first World Bank Chief Economist was appointed, multiple reforms revamped the way analytical work was funded, how it was conducted and reviewed, and how it was connected to lending (Dethier 2009). Even major reorganizational reforms were justified on the need “to improve knowledge flows and the delivery of multisector solutions to clients” (World Bank 2014). However, outsiders would find it difficult to pinpoint any turning point, from an organizational

point of view, that led to a substantial improvement or deterioration in the quality and relevance of World Bank “knowledge”.

At the same time, the quality of the World Bank’s knowledge products is heterogeneous – and at times directly controversial. For example, the rankings and interpretations of the highly visible *Doing Business* report were the subject of much criticism, leading to its thorough reassessment (Manuel et al. 2013). After almost two decades, its production was discontinued on the grounds that “the current methodology should be significantly modified, implying a major overhaul of the project” (Alfaro et al, 2021).

This ambivalence affects less-renowned knowledge products as well. Both insiders and outsiders would readily admit that many books, reports, and articles by the World Bank fail to get traction with academic audiences, policy makers or public opinion. Some would easily qualify as advocacy, which reduces their credibility as a source of reliable knowledge. Others are simply shallow. Such heterogeneity within the same institution raises questions about the processes through which the World Bank’s knowledge-related undertakings are selected, conducted and vetted.

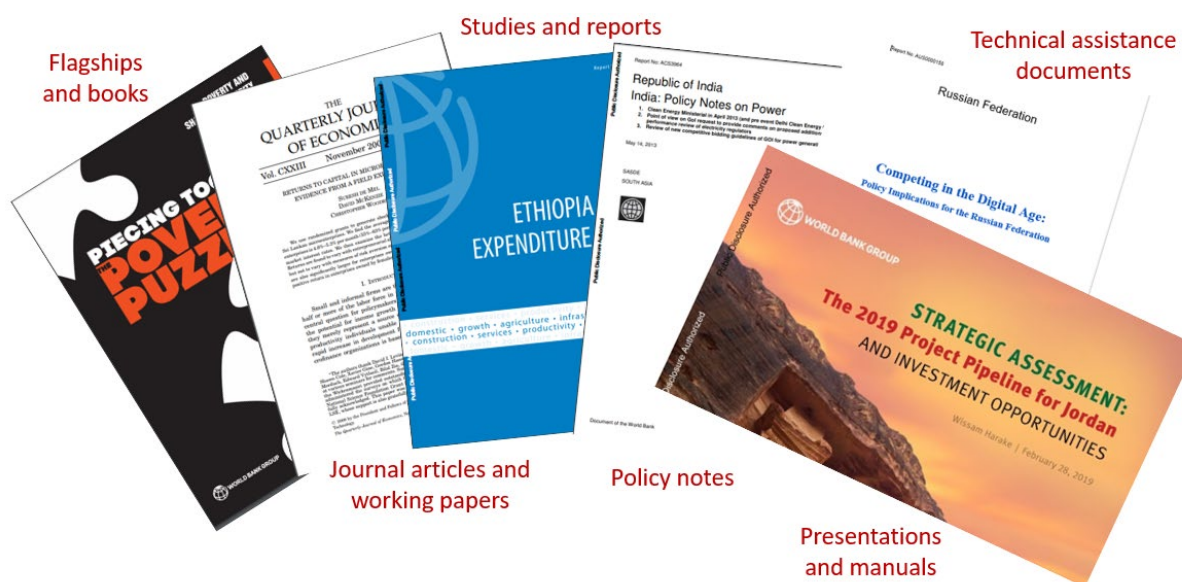
The goal of this paper is to use an empirical approach to shed light on what makes for “good” knowledge work at the World Bank. With the organization being such a prolific source of analytical products, it should be possible to rely on statistical methods to categorize them, to identify key correlates of their quality or impact, and to pinpoint individual items with higher and lower quality within each of the categories.

The fundamental assumption of the paper is that at the World Bank ideas materialize under the form of documents – books, reports, journal articles, working papers, policy notes, presentations, manuals, and the like (figure 1). Given how large and complex the organization is, these documents are by nature addressed to diverse audiences and serve different objectives. Some intend to stimulate debates on economic and social development in general, while others aim to shed light on pressing challenges confronting a specific country or region. Some build on innovative methodologies to push the research frontier, while others digest existing knowledge to build capacity internally or externally. Some are neutral resources whose implications are to be determined by the user, while others encourage the adoption of concrete strategies and policies.

If all knowledge-related documents produced by the World Bank could be classified in relatively homogeneous categories – say, based on their audience, or their purpose – it would in principle be possible to analyze them using a production function approach. This would require some rating of their

success within their category – for example, their technical quality or their communication impact. It would also require information on the process through which they were prepared, from the use of resources to the allocation of responsibilities. If such quantitative approach could be implemented, the key factors most often associated with success could be statistically identified, and outliers – both individually and at relevant levels of aggregation – could be singled out.

Figure 1: Knowledge work at the World Bank is reflected in documents



Previous studies on knowledge products at the World Bank provide guidance on the nature of the output and input indicators to consider in such exercise. However, due to their diverse audiences and objectives, those studies tend to be partial. Some focus on research products that are academic in nature, while others assess dissemination products and operational reports. Some flag the competence of those in the lead, while others emphasize the resources spent. Some discuss the credentials of the units in charge, while others highlight their proximity to end-users.

From a methodological point of view, this paper aims to cover the entire universe of knowledge products generated by the World Bank while at the same time encompassing the various correlates of success explored by previous studies. By combining their different perspectives, the quantitative analysis can be agnostic on what matters the most within each category of knowledge products.

Implementing the proposed approach requires reliable data. Unfortunately, a consolidated database of World Bank knowledge products is not readily available. Systems and platforms have progressed more in

response to budget processes and managerial needs than in connection with the analytical agenda. As a result, the relevant information is currently scattered across the World Bank.

A contribution of this paper is to painstakingly match entries from three self-standing digital platforms to assemble a consolidated database of knowledge-related World Bank work. One of the platforms is a filing system for documents of all sorts, another one contains information on each of the budget codes the World Bank uses, and the third one hosts all the publications it has put in the public domain.

For each of the pieces of knowledge work identified through this matching process, the consolidated database brings together the information available on a broad set of indicators. On the output side, it may report the internal quality rating at completion and the number of downloads over time. On the input side, it may include the name of the task team leader (TTL), the financial resources and time devoted to the task, and the identity of the units in charge of producing it and supervising it. Important gaps remain, as information is not available on all indicators for all entries, but the result is probably the most comprehensive database of World Bank analytical work produced to date.

The analysis in the paper is conducted for knowledge-related documents completed between July 1st, 2015, and June 30, 2019. This choice is determined by the stability of the rules guiding the preparation of analytical work during that four-year period. Indeed, most knowledge-related tasks in the World Bank were by then under the management of central units. These included the research department – or DEC, for Development Economics – as in the past, but also 14 Global Practices introduced on July 1, 2014, as part of a major institutional reorganization. A significant departure from this centralized approach to knowledge generation was six relatively small regional Chief Economist offices, which continued preparing analytical products for the country groupings they served.

This period was also characterized by a relative stability of other relevant criteria, such as the administrative classification of tasks, the identification of “core” knowledge products that all country units are supposed to deliver, the typical budget envelopes by task, and the rules guiding the use of trust funds (TFs). The stability of the context over the four years considered reduces the risk that statistical results may be biased by structural breaks. It also helps address concerns about reverse causality, whereby similar tasks could have been subject to idiosyncratic processes.

The paper proceeds as follows. The next section reviews selected assessments of the knowledge work conducted by the World Bank, summarizing the methodologies they used, and the hypotheses articulated

in each case. Section 3 describes the way the database of knowledge products was constructed and how the key attributes of each document were defined. Descriptive statistics from this database are used in Section 4 to show that the universe of knowledge products is broader than that typically receiving the attention of World Bank management, with both sets even displaying opposite trajectories over time. In section 5, knowledge products are classified in several relatively homogeneous categories; for each of them a statistical analysis akin to the estimation of production functions is conducted. The main patterns emerging from these exercises are summarized in Section 6. Building on them, Section 7 concludes by offering several practical recommendations to strengthen analytical work at the World Bank.

2. Previous studies

Based on their unit of observation, previous assessments of knowledge work at the World Bank can be classified in four groups. A few of them have focused on *research products*, the more academic end of the spectrum. Others, more operational in nature, have tried to distill patterns from *budget codes* that are deemed to generate analytical content. Several studies have put emphasis on the *work processes* through which analytical products are prepared. And a couple of them have elicited and analyzed *user views* in relation to World Bank knowledge products.

Evaluations of *research products* tend to focus on academic recognition, with citations being an obvious measure of the World Bank's influence on development thinking. One study in this group applied the citations metric to some 8,100 journal articles, around 6,800 books and chapters in books, and close to 3,000 working papers that had been put in the public domain by the World Bank over several decades. The set was assembled by combining records from an internal database of documents and from various external bibliographical repositories (Ravallion and Wagstaff 2011).

Using citation data from the Google Scholar platform, this study revealed the heterogeneous quality of the research products selected. A large share of them – including several *World Development Reports* – had little discernable influence. However, within specific geographic areas such as Africa, China and India, and specific topics such as education, growth, inequality, health, and poverty, World Bank research products were ranked highly. And a few of them were among the most cited articles in several of the top economics journals. However, beyond noting that ratings tend to rise with years of publishing, this study did not attempt to explain what drove these stark differences in visibility.

Because it takes years for research to be published, and only then do citations start to accumulate, another study in this first group, conducted by a group of prestigious academics, chose to evaluate the quality of a sample of research products directly (Banerjee et al. 2006). The universe in this case were the journal articles, books, working papers and flagship reports published between 1998 and 2005. In all, roughly 4,000 research products were considered, of which more than 2,000 were articles in peer-reviewed journals. The subset directly evaluated by this study comprised 192 of these products – 180 drawn through stratified random sampling, all the *World Development Reports*, most of the World Bank's flagship reports, and a list of “must read” papers and books.

The study found that many of the documents evaluated met “the highest academic standards of originality and technique” and dealt with topics that were highly relevant to the World Bank's mission. At the same time, however, “much of the research read by the evaluators was seen as undistinguished, and not well-addressed to any particular audience” (Banerjee et al. 2006, p. 46 and 66 respectively). This strong heterogeneity is partly attributed by the authors of the study to the competence of the teams in charge, with a dozen researchers even praised by name.

When reviewing the performance of *World Development Reports*, the study also comments on a “committee process” potentially affecting the quality of knowledge work at the World Bank. This process “makes it difficult to maintain a coherent and focused argument, especially for controversial topics [...] There is also a tendency to pull political punches [...] There is much political correctness, including mindless cheerleading for cultural touchstones such as women, trees, and social capital [...] Trade-offs tend to be eschewed in favor of ubiquitous “win-win” scenarios” (Banerjee et al. 2006, p. 80-81).

Beyond authorship, the study considered other characteristics of the 192 knowledge products evaluated. A simple statistical analysis revealed that overall quality ratings were virtually identical for research products by DEC and by the rest of the World Bank. DEC was clearly stronger on statistical and econometric methodology and in data handling, whereas the rest of the World Bank did better on clarity of writing and appropriateness of recommendations. Outside DEC, flagship reports produced by regional and sectoral units fared very well, and most of the rest quite poorly. And across the entire sample, ratings were lower for products whose preparation involved small budgets, but also for those with very large budgets (Banerjee et al. 2006, Chapter 3, Annex 2).

Studies in the second group focus on *budget codes* that are supposed to be analytical by nature. Chief among them are codes such as Economic and Sector Work (ESW) and Technical Assistance (TA). Together

with codes used for impact evaluation, external training, and other non-lending tasks, ESW and TA are part of what is known in World Bank jargon as Analytical Services and Advisory (ASA) work. Some of these codes, analytical in nature, support the generation of new knowledge on global issues and the discussion of regional and country policy issues. Others, more hands-on, fund technical inputs for developing country governments and capacity building efforts.

World Bank systems record relevant information about ASA budget codes. On the output side, a rating is reported at completion to judge the performance of the task against the objectives stated at its inception. The rating is proposed by the management of the unit responsible for the task, but adjusted – often downward – by the management of the unit that requested it and funded it. For a few of the budget codes, feedback from operational clients is also sought within the year that follows completion. However, for confidentiality reasons, information of this client feedback is available only with some level of aggregation, not for individual budget codes. On the input side, information is generally reported about spending, time to completion, the name of the TTL, the identity of the unit responsible for the product, and that of the requesting unit, among others.

One of the studies in this second group considered 1,331 ASA budget codes deemed completed between 2006 and 2012 (Doemeland and Trevino 2014). The assessment of these codes relied on the consolidated number of downloads for publications associated with them, although no correction was made for the size of the intended audience – for example, whether the budget code concerned a large region or a small island nation.

With this limitation in mind, the study showed that budget codes whose key reports had been released in English, and had been in the public domain for longer, had a higher number of downloads. Importantly, the distribution of downloads was highly skewed, with only 13 percent of the budget codes leading to more than 250 downloads, whereas a full 31 percent had none.

Studies tapping other information available about these budget codes tend to be more informal in nature. They are carried out mostly as part of regular operational management, with some of them covering the entire range of active ASA codes and others referring to a particular region or a particular sector. On occasion, however, some of these assessments go beyond monitoring and try to draw broader lessons for the knowledge work of the World Bank. For example, toward the middle of the period covered by this paper, one such study reviewed how ASA had evolved between 2003 and 2017 (World Bank 2018).

This study found that the number of ASA codes completed annually had grown by 63 percent over the period considered. The composition of the codes had changed as well, with budget codes associated with analytical work declining by a quarter and those supporting advisory activities tripling. The decline was noticeable for the number of budget codes related to “core” analytical products – country economic memorandums, poverty assessments and the like.

The upward trend in budget codes and the change in their composition were attributed by the study to the growing availability of TFs, which are often used to support advisory work. While the average funding provided by the World Bank for a completed task had increased from USD 138,000 to USD 159,000 over the 15 years considered, the study noted that TFs had allowed average lifetime spending per task to jump from USD 208,000 to USD 515,000.

This study also assessed performance at the aggregate level, using the client feedback at completion. It noted that ASA clients valued the World Bank’s ability to bring global expertise but were less impressed by its tailoring to the local context, and by the timeliness of its delivery. However, the study did not attempt to link the performance of individual ASA codes to their characteristics, including total funding or the share of expenditures paid for by TFs.

Studies of World Bank knowledge activities based on *work processes*, the third group of studies takes a more holistic approach, emphasizing the way internal units and external clients interact. Several such assessments have been conducted by the Independent Evaluation Group (IEG), a unit tasked with providing evidence on the delivery of World Bank services and results to its Board of Directors.

An example of this holistic approach is an evaluation of knowledge-related activities in nine country programs (IEG 2016). These were identified among 48 that make relatively intensive use of the World Bank’s analytical work. The selection was not driven by random sampling but rather by an attempt to illustrate a range of engagement modalities at the country level. Taken together, the nine programs covered 266 ASA budget codes and 34 advisory services provided by the International Finance Corporation (IFC), the private sector arm of the World Bank.

Knowledge work in the nine selected countries was evaluated against four criteria: relevance for the client country, technical quality, results achieved, and sustainability of the results. Ratings on those four fronts were based on internal interviews, stakeholder feedback and desk reviews.

This study found that performance was not correlated with budget resources, or with the extent of reliance on TFs. Rather, knowledge products requested by the client and designed specifically to achieve client objectives were more likely to achieve their objectives than those of a more generic character. Knowledge services with fully or partly achieved outcomes also used local expertise, focused on specific sectors rather than broad topics, customized the analysis to local conditions, generated new data in support of policy making, and formulated actionable recommendations. Therefore, the “how” appeared to be more relevant than the “what” to the authors of this study.

The fourth and last group of studies involves conducting surveys to elicit *user views* on World Bank knowledge products. Depending on the case, the users considered may be operational World Bank staff – for example, TTLs of lending projects – or external stakeholders – including government officials, members of legislatures, civil society, the private sector, academics, and donor representatives.

For example, one study by IEG (2009) sought feedback on the usefulness of 129 ESW and 64 TA budget codes that were deemed in line with the overall distribution of knowledge products delivered by the World Bank between 2000 and 2006. Responses were obtained from 91 TTLs of World Bank lending operations who had not been associated with the ESW and TA codes covered – roughly half of the TTLs approached – and from 353 in-country stakeholders.

The study noted that responses were generally consistent between the two groups. Technical quality received high ratings and dissemination less so, resulting in a mixed assessment of overall quality. ESW products such as global and regional reports were said to have informed lending operations, even when that was not their explicit objective. Loans that had been preceded by ESW products were also said to have better design than those that had not. Lending thus appeared as an important channel through which knowledge products informed developed policies and strengthened institutions at the country level. The effect of ESW products on informing World Bank strategies was significant as well, with “core” analytical products being particularly important in this respect.

In another study, a survey of senior World Bank operational staff was used to assess their awareness about research products, such as those produced by DEC (Ravallion and Wagstaff 2011). There were 2,900 recipients of the anonymous survey instrument used, and 555 respondents. Views were similar across regional units, but markedly different across sectors, with staff in areas of economic management and human development being more familiar with research products than their colleagues elsewhere.

However, with the World Bank's research products also being concentrated in those two sectors, it is difficult to tell whether differences in staff awareness really reflect a respondent bias.

On occasion, the relationship between survey responses and features of the World Bank program in a country, including its knowledge work, has been explored through regression analysis. Thus, a study took as its explained variable the ratings on the agenda-setting influence of the World Bank provided by 1,244 government officials from 121 low- and middle-income countries (Knack et al. 2020). This sample was drawn from a list of 6,371 respondents among 54,990 stakeholders in 126 countries, by retaining only public sector stakeholders who had a direct interaction with the World Bank between 2004 and 2013.

Explanatory variables were organized around country-policy clusters – for example, Madagascar and human development. The variables of interest in the regression, all measured at the cluster level, were the number of ESW activities completed during the period of interest, the number of development policy loans approved, and the number of prior actions these loans covered. A prior action documents a supposedly meaningful policy decision adopted by the government before the approval of the loan – for instance, a reduction in trade tariffs. The regression also controlled for stable differences in views associated with these clusters (fixed effects), for annual disbursements per capita from World Bank policy loans and investment projects, for other country characteristics such as population and income per capita, and for individual characteristics of the respondent including gender and age.

In the eyes of the respondents, the results show, the number of ESW activities completed and the number of policy loans approved had a significant influence on the reform activities undertaken by the recipient country. Actual disbursements from policy loans, on the other hand, might have undermined reform impetus, while the number of prior actions was statistically insignificant. The study concludes that analytical products from the World Bank are effective at influencing the design, direction, and implementation of government policy.

Studies in the four groups are necessarily partial, as they attempt to assess the analytical work of the World Bank from one angle, using only a subset of the potentially relevant information. They tend to deal with a selected but potentially narrow set of knowledge products. They often evaluate output quality, but some do so at the aggregate level while others make comparisons across the output set considered. And they do not always link the quality of knowledge work to the underlying production process but they when they do, they tend to emphasize diverse aspects of it.

Yet, when considered together, the studies in the four groups provide a series of *working hypotheses* that are worth exploring more thoroughly. First, the overall quality of World Bank knowledge work is high, but it is heterogeneous both between and within groups of products. Second, the volume of resources available helps, but only to a point, and the growing availability of TFs may in fact reduce overall quality. Third, timeliness is a weak aspect of World Bank analytical work, without indications that knowledge products that take longer to prepare are of better quality. Fourth, the competence of the teams in charge of producing knowledge work matters for its quality, but it can be undermined by management processes that attend to other objectives. Fifth, within the World Bank, DEC is stronger on technical rigor, but possibly weaker on communication clarity and policy relevance. And sixth, the overall quality of World Bank knowledge work is higher when it is conducted in greater proximity to its intended users, hence with a better understanding of their needs and context.

3. A database of knowledge products

An empirical assessment of the six working hypotheses drawn from previous studies requires a database of knowledge products that is comprehensive both in its coverage and in the information set available for each entry. Specifically, data is needed on all the analytical work produced by the World Bank during a specific period, on the quality and reach of each piece, on the competence of the team in charge, on the budget resources devoted to it, on the time it took to complete it, on whether DEC was involved, and on the proximity between the preparation process and the ultimate client.

A comprehensive database along those lines was assembled for this paper by matching information available in three independent digital platforms:

- The *Image Bank* (IB) is a document filing system where all entries are classified based on their sensitivity, which determines who inside or outside the World Bank can retrieve them. The criteria to decide which documents are filed have some fluidity; only a fraction of them qualify as knowledge products.
- The *Business Intelligence Warehouse* (WH) is a digital platform organized around budget codes that is accessible only internally. A budget code supports the production of multiple documents, but there is typically one document summarizing its output – for example, a project appraisal document or a technical assistance report. Analytical work is supposed to be funded through ASA budget codes.

- Finally, the *Open Knowledge Repository* (OKR) makes all World Bank publications available to the public at large. Long-running periodical series and the annual workplans of various World Bank units determine which documents are to be disseminated through this repository. However, as the name of the database indicates, all the documents in it can be deemed related to knowledge.

The matching exercise was conducted for documents filed in the IB database between July 1, 2015, and June 30, 2019, a period that corresponds to the World Bank Fiscal Years 2016 to 2019, or FY16-FY19 for short. In the case of the WH database, the matching concerned all budget codes that had been deemed completed in FY16-FY19, plus all active codes as of June 30, 2019. Including the latter group is justified because documents may be filed in the IB database before the corresponding budget codes are closed. Similarly, to account for delays linked to administrative clearances and production processes, all publications released through the OKR database after July 1, 2015, were included in the exercise.

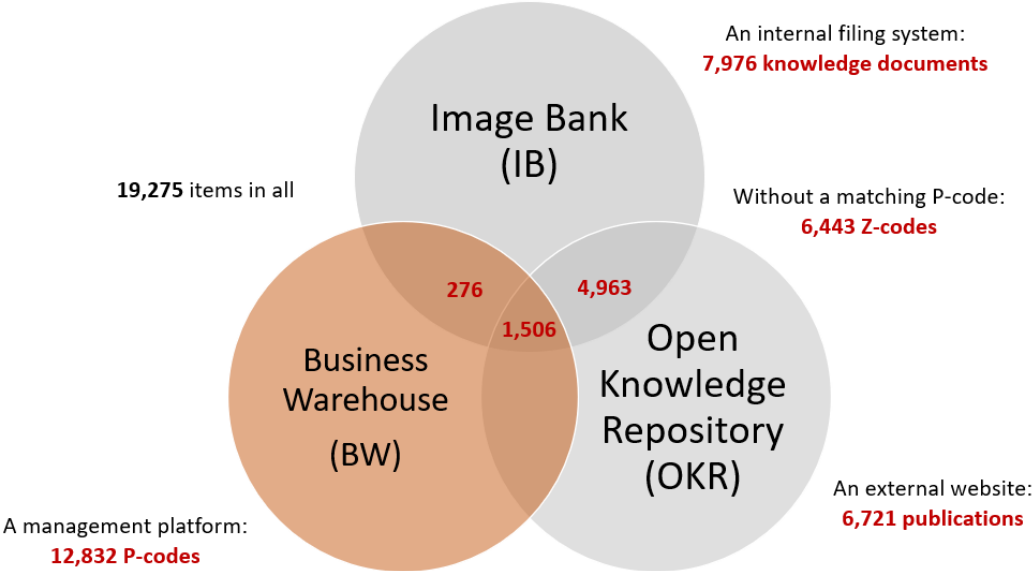
An important challenge was to determine which of the documents in the IB database had substantial knowledge content. Documents are filed there under one of more than two dozen headings, such as Country Economic Memorandums, or Policy Notes, or Sector Strategies (table A.1). However, several thousand documents were not tagged to any of these ten categories and needed to be individually checked to determine whether they qualified as knowledge products.

Similarly, only a fraction of the budget codes in the WH database support knowledge work. ASA codes in principle do, so they were all included in the matching exercise. But knowledge work is also supported by budget codes funding DEC research, as well as by the preparation of regional and sectoral strategies, country partnership frameworks, and training, among others. Conversely, a considerable amount of knowledge work is produced without mobilizing earmarked budget resources.

The matching process focused first on budget codes, which are at times mentioned in the documents filed in the IB database and in the publications released through the OKR database. Successful matches were indexed by the number of the corresponding budget codes in the WH database, known as P-codes in World Bank parlance. When no P-code could be identified in IB documents or OKR publications, the matching process relied on information about the titles of the entries and the names of their authors. Further confirmation that the matching was correct was provided by the regions or countries these entries covered and the sectors they were mapped to. Entries without associated P-codes were indexed with numbers allocated sequentially; these numbers are identified in what follows as Z-codes.

In all, the consolidated database used in this paper contains 12,832 P-codes and 6,443 Z-codes, jointly representing the entire set of knowledge products generated by the World Bank in FY16-FY19. Significant data gaps imply that only 1,506 of these products can be found simultaneously in the IB, WH and OKR databases. Another 5,239 knowledge products appeared in two of the databases, and the remaining 12,530 in only one of them (figure 2).

Figure 2: Three partially overlapping databases of knowledge work



Combining information from the three databases, data on up to 14 variables was obtained for each of the 19,272 knowledge entries in the consolidated database. On the output side, a key variable was the rating given to the task by the unit requesting and funding it, relative to the objectives set at its inception; another one was the total number of downloads across platforms. On the input side, there was information on the World Bank resources and TF money that were spent during the preparation of the task, on the time it took to complete it, on the name of the TTL or key author, and on those of the units in charge of delivering the product and of supervising and approving it.

Several additional adjustments were made to these variables:

- Whenever available, downloads from the IB and OKR databases were added. The total was scaled by the population of the relevant region or country in 2019, using demographic data from the United Nations (2020). Results were expressed in downloads per million people.

- Task ratings in the WH database were consolidated into three categories. High (= 1) corresponds to “extremely effective”, “highly effective” and “very effective” ratings. Medium (= 2/3) to “satisfactory”, “moderately satisfactory”, “effective” and “moderately effective”. Low (= 1/3) is for the rest.
- Google Scholar, Research Papers in Economics, and the Social Science Research Network were used to assess the technical skill of TTLs. Those who showed up in at least two of these platforms were deemed Top (= 1). Those in one only were considered Good (= 2/3), and the rest Unknown (= 1/3).
- The numerous units in charge of delivering knowledge products were consolidated into four groupings: Global Practices – which also includes central Cross-Cutting Solutions teams dealing with topics such as gender, or climate change – DEC, regional Chief Economist offices, and the rest.

The consolidated database of knowledge products assembled through this process is available on request.

4. Composition and summary statistics

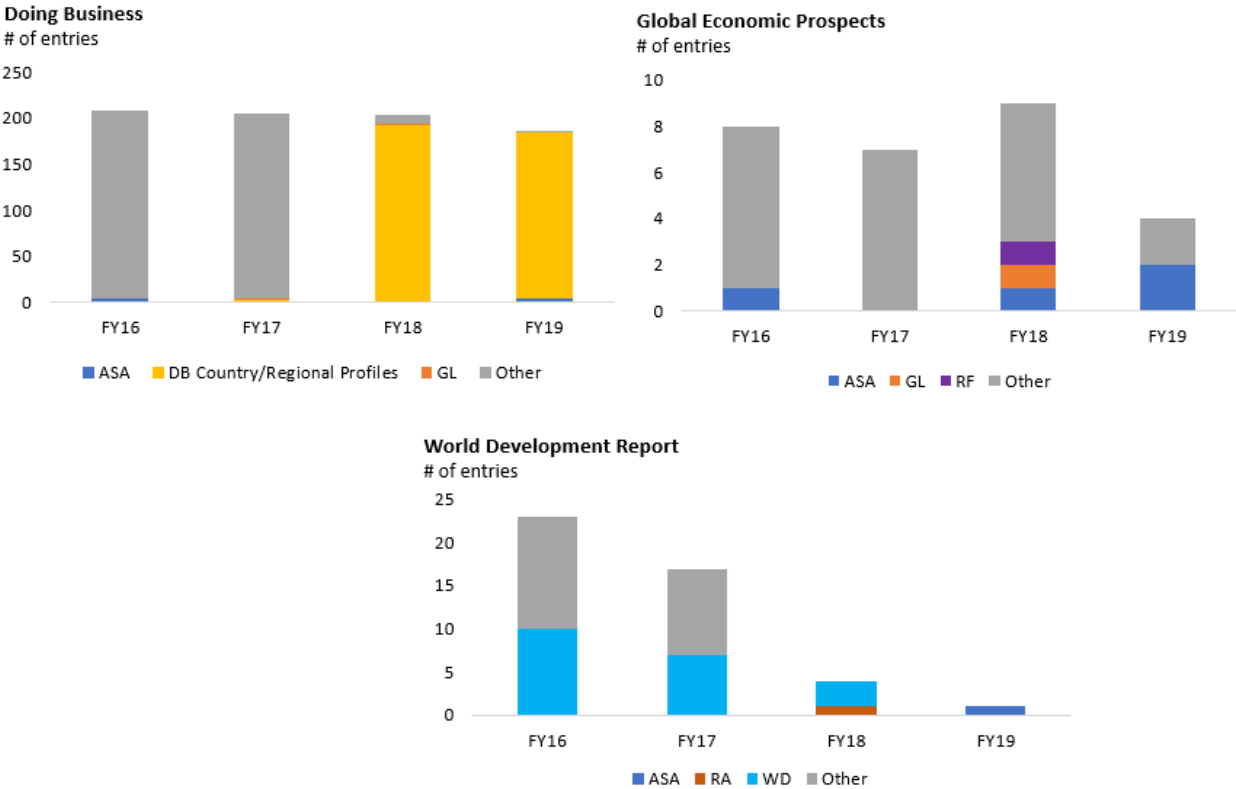
While the consolidated database is structured around P-codes and Z-codes, this may not be the most informative breakdown to assess the quality of the analytical products in it. Admittedly, knowledge-related activities are monitored by World Bank based on ASA budget codes, which are a subset of P-codes. Quite often their oversight involves counting how much was spent on them, and how many activities were completed over a specific period. This approach allows disciplining the use of budget resources and ensuring the delivery of committed products.

As for the quality of the analytical work, the conventional wisdom is that it is inversely correlated to the number of tasks management needs to oversee. A new standard monitoring report for ASA products introduced at the beginning of the period covered by this study was used to drive a substantial “cleanup” of analytical products (World Bank 2019). Since then, considerable managerial effort has gone into ensuring that the number of active ASA products is commensurate with some experience-based indicator of the “capacity” of responsible units to deliver.

However, the number of active ASA budget codes may not provide a reliable assessment of the number of knowledge-related tasks being conducted, due to the significant heterogeneity in coding practices across World Bank units. Consider *Doing Business*, *Global Economic Prospects*, and the *World Development Report* – three of the most visible analytical products of the World Bank. Depending on the

fiscal year, their preparation was associated with zero, one or several ASA codes each. At the same time, these reports mobilized non-ASA codes, in varying numbers. For example, there were close to 200 additional codes associated with each edition *Doing Business*, with the main report and its 11 regional profiles having their own P-codes, whereas country profiles – which are subject to a less structured review process – often appeared as Z-codes. On the other hand, there were only a few extra codes associated with each edition of the *World Development Report*, and none in FY16 (figure 3).

Figure 3. Knowledge products of similar importance are coded in different ways



As important as ASA budget codes are, the consolidated database shows that they accounted for about 55 percent of the knowledge-related work delivered by the World Bank in FY16-FY19. An additional 4 percent was from budget codes supporting research activities by DEC, and another 8 percent from various other budget codes. The remaining third of the analytical products delivered had no budget code associated with it (table 1).

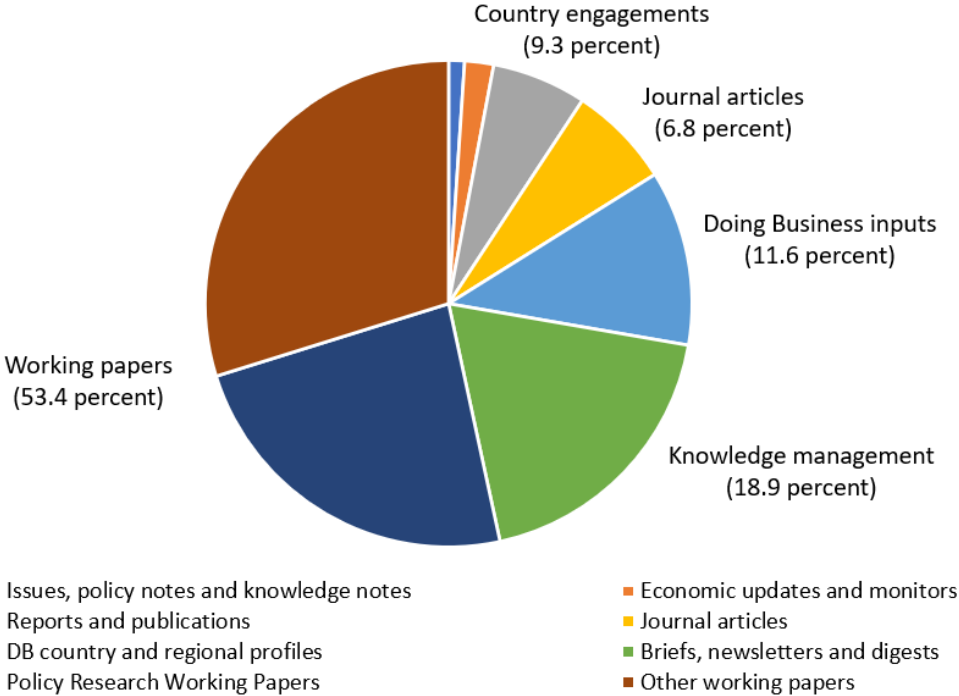
Many of the analytical products not supported by ASA budget codes are arguably significant from a knowledge point of view. This is almost certainly the case for the research products generated by DEC. It

may also be so for a large fraction of the Z-codes in the consolidated database. Almost 7 percent of these correspond to journal articles, and another 53 percent to working papers, two types of knowledge products that greatly contribute to the academic standing of the World Bank. Also, about 9 percent of Z-codes correspond to country engagements such as economic updates and policy notes, which tend to be among the analytical inputs most valued by local counterparts (figure 4).

Table 1. Budget codes covered by the consolidated database

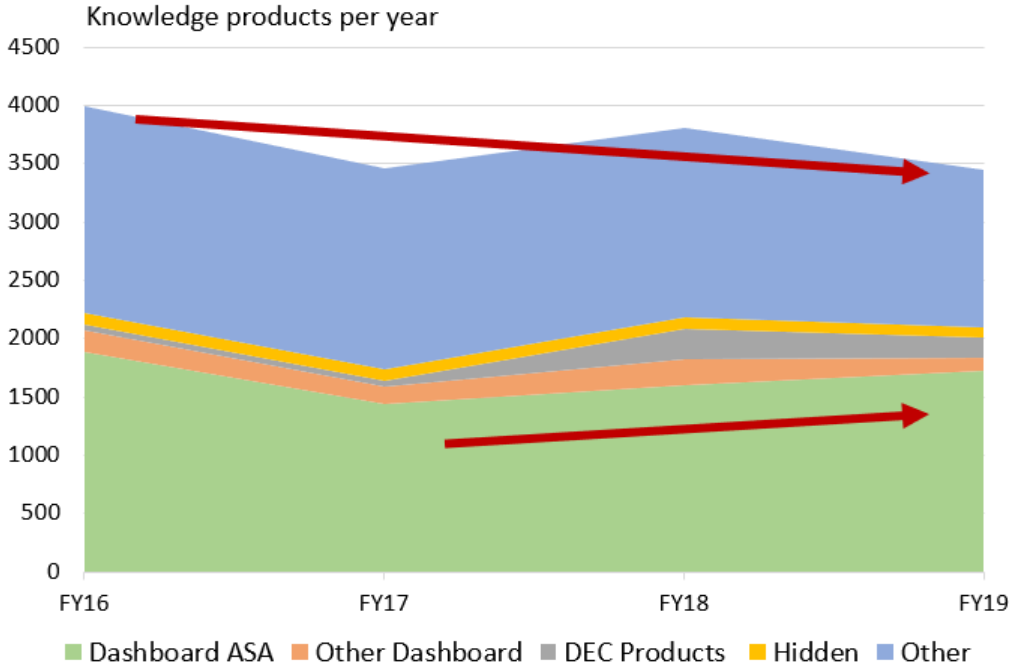
ASA			Hidden		
AA	Analytical and Advisory	6675	CP	Country Partnership	253
EW	Economic Sector Work	1045	PE	IDA Guarantee	76
IE	Impact Evaluation	116	ES/EP/ER/EC	Evaluations	30
PA	Prognostic Approach	318	RE	Recipient Executed Activity	10
TA	Technical Assistance	2333	PT	Partnership	8
TE	Client Training	144	CN/GE	Carbon offset/ GEF	6
DEC Products			GP	Global Partnerships	5
RA	Research Services	189	SP	Sector strategy	2
RF	Research Services	536	Other		
Other Dashboard			Z-code		6443
GL	Global Monitoring	130			
KP	Knowledge Products	953			
WD	World Development Report	3			

Figure 4. Many significant knowledge products do not have an associated budget code



Finally, it is worth noting that the number of knowledge products delivered in connection with ASA budget codes may not vary over time in the same direction as the overall number of analytical products generated by the World Bank (figure 5). The annual number of ASA-funded analytical products was roughly the same in FY16 and FY19; it even increased between FY17 and FY19, despite the efforts to keep knowledge work within management’s oversight. However, the actual number of analytical products delivered declined from roughly 4,000 in FY16 to less than 3,500 in FY19.

Figure 5. Budget codes provide a partial measure of the World Bank’s knowledge work



A potentially more informative classification of knowledge products considers not the budget processes supporting their production, but rather the objectives they serve, their audience (global, regional, or national), their expected outcomes and the risks they pose. While retaining a manageable number of categories, the proposed tiering aims to achieve a greater homogeneity of products within each of them. This should allow more meaningful quality comparisons across entries. It should also make it possible to rely on a production function approach to identify the key correlates of higher quality for each tier.

Seven tiers of knowledge products are retained in what follows (table 2). The consolidated database contains information on output and input indicators by product across most categories, but data coverage tends to be weaker in the lower tiers of the proposed classification (table 3).

Table 2. A functional tiering of World Bank knowledge products

Tier	Knowledge products included	P-code	Z-code
Flagships	World Development Report; Doing Business; Global Economic Prospects; regional and sectoral flagships.	132	52
Debates	Regional, sectoral and multi-country reports; Doing Business regional reports.	2219	628
Extended core	Country Economic Memorandums; Poverty Assessments; Systematic Country Diagnostics; Public Expenditure Reviews; Reimbursable Advisory Services; Financial Sector Assessment Programs; Risk and Resilience Assessments; Country Private Sector Diagnostics; Human Capital Reports; Infrastructure Sector Assessment Programs; Debt Management Performance Assessments; Debt Sustainability Analyses; Agricultural Sector Reviews; Pandemic Preparedness Reports; Environmental Assessments.	1500	6
Other country	Policy Notes; Economic Updates; other studies; technical assistance.	6103	1258
Research	Journal articles; working papers; impact evaluations; Doing Business country profiles.	1725	3927
Capacity	Toolkits; databases; conferences; courses; training.	488	82
Internal	International Development Agency papers; strategy documents; briefs; newsletters; evaluations.	665	490

5. Regression analysis

The construction of a consolidated database of knowledge products containing tens of thousands of observations, each with information on its performance and the process through which it was prepared, makes it possible to assess what makes for “good” analytical work at the World Bank using statistical tools. The assessment is in the spirit of a production function, with a measure of performance as the explained variable, and a series of “inputs” to the preparation of knowledge products as the explanatory variables.

The quality of World Bank knowledge products is judged from two complementary perspectives. The rating given to a task by the management of its requesting unit indicates the extent to which it attained the objectives set at its inception, hence of its attainment of expected standards. The number of downloads per million population in the countries or regions that were the focus of the task provides an indication of the visibility it attained, hence of its potential development impact.

Table 3. Summary statistics by tier of knowledge products

Indicators		Flagships	Debates	Extended core	Other country	Research	Capacity building	Internal
<i>Rating</i>	High (%)	11.4	13.3	16.4	14.0	7.5	10.7	2.0
	Medium (%)	0.0	0.9	0.7	1.0	0.5	0.4	0.3
	Low (%)	11.4	28.1	28.5	41.5	21.0	19.3	3.2
	Unknown (%)	77.3	57.6	54.4	43.5	71.1	69.7	94.6
<i>Downloads</i>	Mean (per million)	613.0	62.4	18.1	19.4	12.7	8.6	5.0
	Unknown (%)	71.2	89.7	59.8	87.2	75.5	95.5	91.0
<i>Cost</i>	World Bank (K\$)	793.8	155.0	152.0	114.5	186.6	156.4	215.0
	Trust fund (K\$)	244.4	413.2	55.2	255.6	373.1	277.4	176.7
	Total cost (K\$)	1064.0	592.2	394.4	383.8	582.1	442.9	406.3
	Unknown (%)	6.1	11.9	8.4	11.1	9.1	15.8	11.7
<i>Time to completion</i>	Mean (months)	54.7	26.0	21.2	21.5	41.4	31.8	31.1
	Unknown (%)	0.8	0.4	0.3	0.3	0.9	0.2	3.8
<i>TTL profile</i>	Top (%)	13.6	6.8	2.3	3.6	26.8	4.9	2.4
	Good (%)	2.3	4.0	5.1	4.3	4.2	1.8	2.9
	Known (%)	4.5	4.4	9.9	5.2	3.6	4.7	7.4
	Unknown (%)	79.5	84.9	82.7	86.8	65.4	88.5	87.4
<i>Responsible unit</i>	DEC (%)	11.7	1.2	0.0	0.4	11.5	8.2	1.2
	Global Practice (%)	45.2	41.8	0.6	0.7	4.0	41.2	46.4
	Regional office (%)	2.7	23.4	83.7	70.1	12.4	23.3	1.5
	Other (%)	12.2	11.6	11.7	11.6	2.7	12.8	11.6
	Unknown (%)	28.2	22.1	0.4	17.1	69.5	14.4	42.7

These two measures focus on different aspects of performance; not surprisingly, their correlation across the entire database is -0.11. The relevance of the two measures also varies across task tiers; for example, a knowledge product prepared for internal use by the World Bank should be expected to meet high standards but not to have a substantial number of downloads. With these caveats in mind, running all production function estimations on both measures should provide more reassurance on the validity of the estimates. If a particular input in the production function appears to significantly impact both ratings and downloads, then it can be safely concluded that it does matter for the overall product quality.

Inputs are selected in a way allowing to evaluate the four working hypotheses articulated after reviewing previous studies on the correlates of good analytical work at the World Bank. A first group of inputs includes the amount of resources devoted to a task, the TF share of those resources, and the time allowed for its preparation. These are the inputs emphasized by studies that take *budget codes* as their unit of observation. Second is the profile of the TTL, a potentially significant quality determinant according to studies focused on *research products*. And third is the type of World Bank unit in charge of preparing the knowledge products which, based on studies dealing with *work processes* and *user views*, could affect the relevance of the tasks and their attention to client needs.

The production function analysis also controls for the fiscal year when a knowledge product was delivered and for the geographic region it covered. This is to account for systematic differences in performance over time and across space. For example, it could well be that the broader context in which the institution operated in a particular fiscal year affected management attention to knowledge products. Or that clients in higher-income regions demand better-quality knowledge products from the World Bank. Different degrees of internet penetration across countries should affect the number of downloads as well.

All regressions start parsimoniously, including only the first group of inputs, then add the TTL profile, and finally the type of unit in charge. Fixed effects for fiscal years are included in all specifications, but estimates are presented with and without fixed effects for geographic regions. This sequential approach allows assessing whether the resources and time devoted to knowledge products are robust determinants of their quality. They would not be if the estimated impact of the first set of inputs mattered in the initial, parsimonious estimation but lost significance when controlling for the other inputs.

Finally, the analysis needs to take into consideration the fundamental heterogeneity of World Bank knowledge products. A crude way to do so is to include all observations in the estimation but to add to the specification dummy variables for the various tiers of knowledge products identified above. Proceeding this way amounts to assuming that inputs have the same impact on performance across all

tiers, but the very nature of the products in each tier systematically affects their ratings, or their number of downloads. For example, the dissemination process for flagships could result, other things equal, in more downloads compared to other knowledge products.

However, it is plausible that the production function of analytical products differs across tiers. For example, a TTL with a strong technical profile may be indispensable in the preparation of a flagship report, less so for other analytical tasks. To account for this possibility, regressions are also run separately for two different subsets of observations. One of them includes all major reports, or flagships and debates in terms of the proposed tiering. The other contains all country-specific knowledge products, corresponding to the extended core and other country tiers. Even so, a dummy variable is included to capture any remaining differences between the two tiers lumped in each of the two subsets of observations.

In all, then, six regression analyses are run. The first three have ratings as the explained variable, but differ in their sample, which sequentially includes all observations with data, only flagships and debates, and only extended core and other country products (tables A.2 to A.4). The other three regressions use the same samples sequentially but have downloads as their explained variable (tables A.5 to A.7).

All regressions are run using ordinary least squares (OLS). The first set of regressions could in principle be estimated using a Logit or a Probit model, given that the explained variable is ordinal, not cardinal. Relying on OLS implicitly assumes that the difference in quality between a high and a medium rating is the same as the difference between a medium and a low rating. However, the apparently greater precision of a Logit or a Probit regression would be misleading. The high, medium and low categories for ratings are themselves built by aggregating more granular qualitative ratings, such as satisfactory, moderately satisfactory, effective and moderately effective. Such aggregation was needed due to the number of options available changing during the period covered by this study. In practice, the results of Logit and Probit regressions are similar to those obtained with OLS, but this illusory extra rigor would come at the expense of the fit of the regressions (results are available on request).

Because information is not available on all correlates for all knowledge products, the number of observations included in the regressions is substantially lower than the total number of knowledge products. Information on cost, use of TF resources and time to completion is only available for P-codes. Because these three correlates are included in all specifications, all regressions exclude Z-codes.

With the regression analyses covering about 60 percent of all knowledge products, results should be interpreted with caution. Indeed, the observations included in the regressions could differ systematically from those excluded, in which case the results may not be extrapolated to the entire sample. For example,

working papers and journal articles – typically recorded as Z-codes – are underrepresented in the regressions. Whether results are valid for the entire category of analytical products considered depends on the similarity – or lack thereof – between included and excluded observations.

6. Main results

The large number of regressions that results from combining two explained variables, six specifications and three samples may somewhat blur the findings. To handle this complexity, summary indicators across all specifications can be constructed for each of the regressions, resulting in a heatmap where color indicates sign and intensity the robustness of the result. Thus, dark green (red) means that the estimate is significantly positive (negative) and of the same sign in at least half of the specifications. A lighter color is used when less than half of the estimates meet the criterion above, and no color at all when all estimates are statistically insignificant, or some of them are significant but with opposite signs. The significance threshold is set at 10 percent, which is relatively low but yields a more granular heatmap (table 4).

Table 4. A summary of results from the regression analyses

Explanatory variables	Ratings			Downloads		
	All knowledge products	Flagships and debates	Country products	All knowledge products	Flagships and debates	Country products
Total cost (K\$)	Dark Green	Dark Green	White	Dark Green	Light Green	White
Trust funds (percent of cost)	Red	Red	Red	Red	Red	White
Time to completion (months)	Red	Red	White	Red	Red	White
Top TTL	Red	Light Red	Red	Dark Green	Light Green	White
Good TTL	Dark Green	Dark Green	Dark Green	White	White	White
DEC	Dark Green	Red	Dark Green	White	White	Red
Global Practice	Dark Green	White	Dark Green	Red	White	Red
Regional CE office	Dark Green	White	Dark Green	Dark Green	Dark Green	Red
FY17	Dark Green	Dark Green	Dark Green	Dark Green	Light Green	White
FY18	Dark Green	Dark Green	Dark Green	Dark Green	White	White
FY19	Dark Green	Dark Green	Dark Green	Dark Green	Dark Green	Red
F-test	Dark Green	Dark Green	Dark Green	Dark Green	Light Green	Dark Green

Note: Dark green (red) indicates that the estimate is significantly positive (negative) and of the same sign in at least half of the specifications. A light color indicates that this criterion is met by less than half. No color is reported when all estimates are statistically insignificant, or some are statistically significant but with opposite sign. The significance threshold is 10 percent.

The overall significance of the regressions, reflected in their F-test, allows assessing the relevance of the production function approach chosen in this paper. The approach is robust to explain variation in ratings, both across the entire sample and for the selected subsets of knowledge products. It can also account for some of the dispersion in downloads across knowledge products, but its explanatory power is lower than for ratings, suggesting that there are other important factors at play, in addition to the correlates identified based on previous studies.

The sign and significance of the coefficients associated with those correlates can in turn be used to assess the working hypotheses that had been derived from previous studies. Starting with funding, across the entire sample there is a statistically significant association between the volume of resources devoted to analytical tasks and both their ratings and their downloads. This result also holds, in the case of flagships and debates, although it is somewhat weaker. On the other hand, the impact of greater resources is statistically insignificant in the case of country products.

For a given cost, performance on both ratings and downloads declines as the share of resources coming from TFs increases. This unambiguous result may be due to management devoting less attention to the way third-party resources are used, relative to its own resources. However, worse performance could also reflect a selection bias, if TFs tended to be allocated to analytical tasks that are less relevant and thus tend to get lower oversight and lower appreciation. Only in the case of country products is reliance on TFs uncorrelated with the number of downloads.

Another key resource is the time available to complete an analytical task. A natural presumption is that a greater time allocation results in a better product. However, the estimated coefficients indicate that the longer an activity takes to be completed, the lower its rating and downloads tend to be. This result is highly significant both across the entire sample and for flagships and debates. However, the time to completion does not seem to affect the performance of country products, suggesting that more idiosyncratic factors are at play in their case.

Expectedly, the profile of the TTL matters for the performance of analytical tasks, but with important nuances. Good TTLs are associated with significantly higher ratings in all the samples considered but not with a larger number of downloads. The opposite is true for more academically oriented TTLs, who receive significantly lower ratings internally, but generate a greater visibility for the World Bank. These results highlight the value of strong technical skills when conducting analytical work, but they also reveal a different comparative advantage across TTLs. Those who have some academic recognition seem better

suited to interact with external audiences, whereas solid but less stellar ones may be more able to deliver the type of analytical products World Bank management seeks.

The nature of the unit in charge of overseeing analytical tasks matters too. Across the entire sample, both ratings and downloads are significantly higher for tasks managed by regional Chief Economist offices. Flagships and debates also get more downloads too, but not necessarily better ratings relative to other units in the World Bank. Performance is definitely more mixed for country products, with tasks managed by regional Chief Economist offices getting significantly higher ratings, but also significantly lower downloads. This contrast could be indicative of findings and recommendations that are not so much appreciated by local audiences.

Results are more mixed in the case of Global Practices. Their flagships and debates clearly do not stand out. However, across both the entire sample and the subsample of country products, the tasks they manage get consistently higher ratings and consistently lower downloads. This contrast suggests a greater ability of Global Practices to cater to the needs of World Bank management than to the curiosity of external audiences.

Perhaps the most puzzling findings concern DEC, which is arguably the World Bank unit with the strongest research credentials. Internally, analytical products managed by DEC get somewhat better ratings than those of other units across both the entire sample and the subsample of country products, but not in the case of flagships and debates. On the other hand, DEC analytical products do not get greater visibility in general, nor in the case of flagships and debates. And their visibility is significantly lower in the case of analytical products for specific countries.

DEC's relatively muted splash with broader audiences, despite its unquestionable analytical firepower, could be a consequence of its unique profile. With its greater focus on the debates that agitate academics, and on the methodologies that push the research frontier, DEC may not be ideally positioned to identify the development challenges country audiences are grappling with, and to deliver the simple but practical messages they are looking for.

In addition to assessing the working hypotheses drawn from previous studies, the regression analyses above shed light on the quality of the World Bank's analytical work across time and space. Starting with time, ratings are significantly higher in FY17-FY19 than in FY16. This result is plausible, as the period considered in this paper immediately followed a major reorganization of the World Bank, one that might have temporarily detracted management attention from the supervision of analytical tasks.

Across the entire sample, the results also show some increase in the number of downloads. But this apparent improvement may simply reflect the progress in internet penetration and a greater savviness of task teams to disseminate their findings online. This upward trend is also visible, to some extent, in the case of flagships and debates. On the other hand, the visibility of analytical country products by the World Bank remains stagnant despite the obvious digital advances, and even diminishes toward the end of the period considered.

The disconnect between improving ratings and stagnant or even declining downloads for country products could reflect a drift toward self-complacency in evaluating analytical tasks, but this is a hypothesis that cannot be evaluated with the information at hand. On the other hand, the regression analyses allow identifying where high ratings coexist with limited downloads, or the other way around. This could be done for individual analytical tasks, by comparing their predicted rating and downloads to the actual values of both indicators. However, margins of error may be too large for the findings to be conclusive, making it preferable to estimate performance gaps at some level of aggregation.

Focusing on geographic areas is potentially informative in this respect, as operational units at the World Bank tend to match sovereign national jurisdictions and country groupings connected by their history. Country management units are also a major source of demand and funding for the institution's analytical work, and the main interface with its intended audience.

How ratings and downloads vary across such units can be assessed by replacing regional dummies by country dummies in the regressions above. For this alternative regression exercise, dummy values are set equal to one if the analytical task focuses on the country or on the broader region the country belongs to; they are set equal to zero otherwise. The most comprehensive specification, including all correlates, is used for this exercise (figure 6).

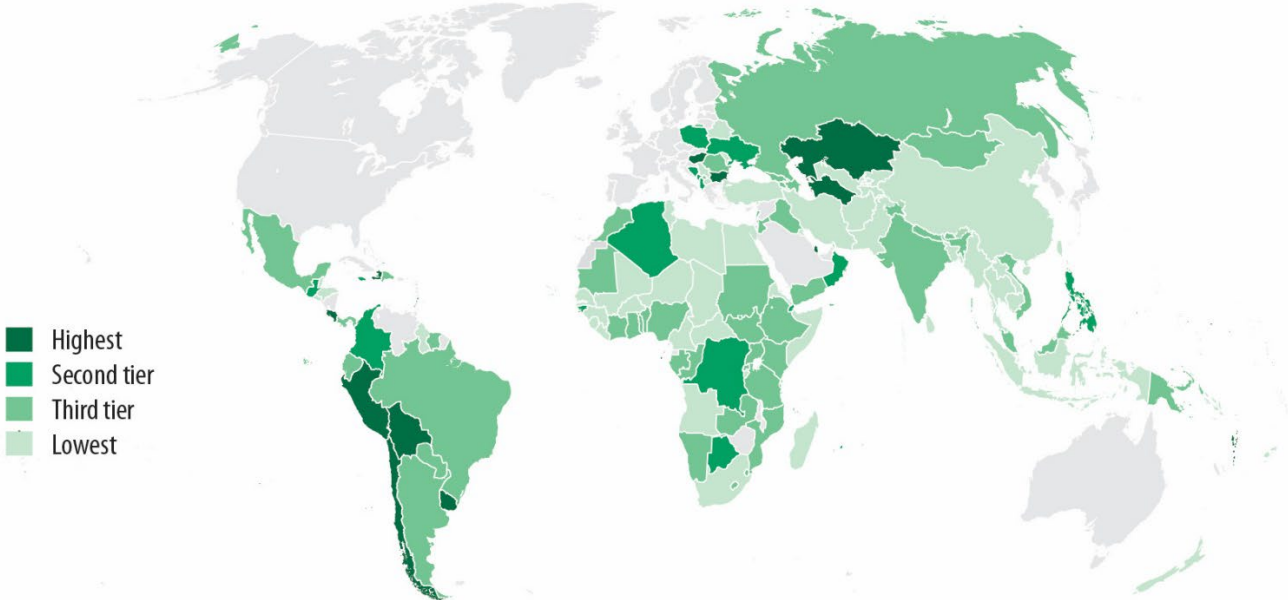
When proceeding this way, the estimated coefficients associated with the correlates resemble those obtained using regional dummies (results are available on request). And the coefficients associated with the country dummies are jointly significant. While relatively few of them are individually significant, these estimates remain the best available predictor of country effects.

Other things equal, ratings were substantially higher in much Latin America and the Caribbean, in several countries in Europe and Central Asia, in parts of Southeast Asia and in most of West Africa. But the geographic distribution of country effects for downloads was quite different. The visibility of the World Bank's analytical work was indeed concentrated in fewer countries, and especially in Bolivia, Ethiopia and

China. A second tier of higher downloads includes the Arab Republic of Egypt, Indonesia, the Islamic Republic of Iran, Iraq, Madagascar, Mexico, Mongolia, South Africa and Türkiye, among others.

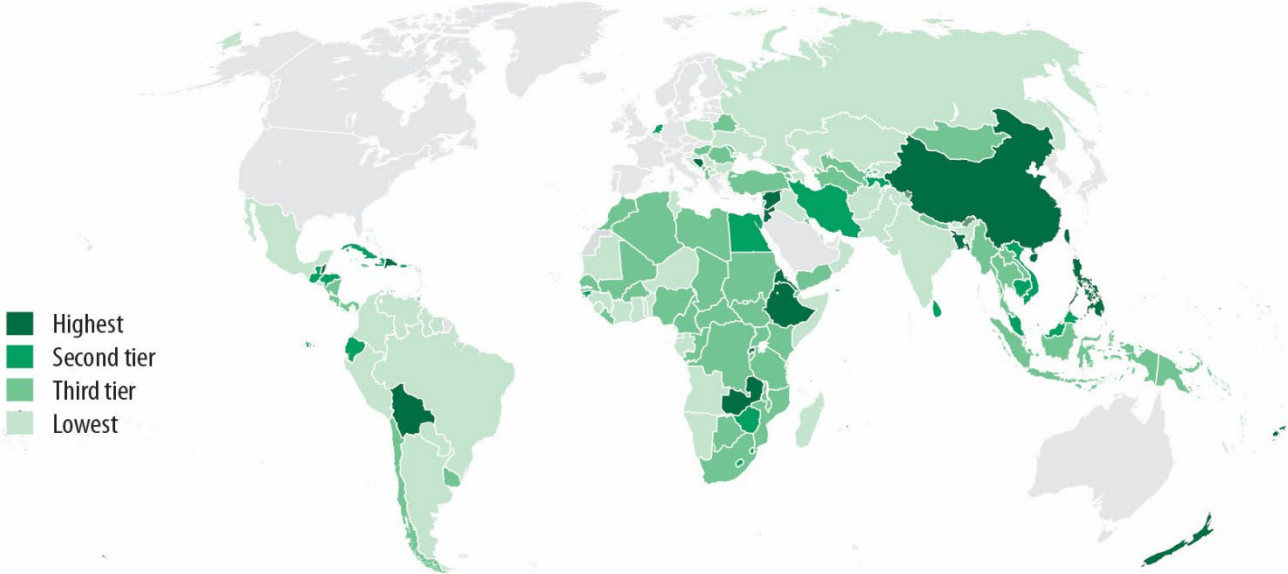
Figure 6. Ratings and downloads differ systematically across countries

a. Country effects on ratings (all analytical tasks)



IBRD 47742 |
JANUARY 2024

b. Country effects on downloads (all analytical tasks)



IBRD 47743 |
JANUARY 2024

7. Conclusions

The analysis in this paper confirms that the World Bank is a prolific producer of analytical work: much more prolific in fact than what its information systems for knowledge management suggest. By focusing on documents, and by matching data about them from three independent digital platforms, this paper shows that the number and scope of analytical products delivered over the four years considered for the analysis vastly exceeded the number of analytical budget codes overseen by the World Bank.

Discrepancies between the two counts are not stable, both throughout the years and across similar products at any point in time. Starting with years, the gap between the number of analytical products delivered and those captured by information systems varied substantially over time. As for products, the number of tasks associated with major analytical undertakings – such as the World Development Report, Doing Business and Global Economic Prospects – was different as well, despite their institutional relevance supposedly being similar.

By focusing on the entire set of knowledge products, and by considering many of their characteristics at once, the paper assessed the validity of several received hypotheses about the drivers of good analytical work at the World Bank. These hypotheses had been formulated by previous studies that relied on different units of observation and focused on partial sets of task characteristics. The findings of the paper validate most of the hypotheses, but the broader dataset used yields empirical regularities that are arguably more precise and robust than those of previous studies. The analysis also leads to novel findings.

Specifically, results show that devoting more resources to an analytical task leads to better ratings but not to a greater number of downloads, and that ratings are systematically worse when resources come from TFs. More time devoted to a task also tends to be associated with poorer ratings, without a noticeable impact on visibility. The paper also finds that the profile of TTLs matters as well, but not in a linear manner. More academically oriented TTLs underperform on both ratings and downloads, whereas technically solid but less stellar TTLs overperform. The pattern is similar in relation to the unit in charge. The more research-oriented DEC underperforms in ratings relative to both Global Practices and regional Chief Economists offices. The latter also outperform other units in terms of downloads.

Some of the results in the paper will hopefully be useful for World Bank management when launching new analytical tasks or monitoring existing ones. They may also be of interest for other development institutions when thinking about their own analytical products.

The paper can also be read as a cautionary note against knowledge management that is based on counting analytical tasks. For quite some time, it has been a common practice at the World Bank to try to align the number of active ASA budget codes with some “capacity indicator” established in a relatively crude way, based on what was delivered in previous years. In an institution like the World Bank there are simply too many analytical products underway at any point in time for this to be an effective approach, especially when a large share of such products is not even “visible” through information systems.

Instead, the methodology and findings of the paper can be seen as a call for a different knowledge management approach, one that relies on stronger information systems, a better alignment of incentives, and regular monitoring and evaluation based on statistical analysis.

Starting with information systems, a single identifier for analytical tasks would allow connecting data on all the resources they use, all the outputs they deliver, all the downloads they receive, and all the feedback from internal and external clients. Budget identifiers (P-codes) are a practical way to connect information that is currently scattered across various disconnected systems. At present, no TF can be created in the World Bank without a link to a P-code. Arguably, it should be the same for every document entered into the IB and OKR digital platforms.

The process could be hardwired, facilitating an automatic linking of the three databases used in this paper. A similar approach could be used to connect budget codes with the user feedback collected from country engagement surveys. Proceeding this way, a time will come when it will be possible to conduct data-driven studies like the one in this paper without having to go through a laborious reconciliation of entries.

On incentives, thousands of analytical tasks running in parallel call for decentralized management. But knowledge products differ in the objectives they serve, their expected outcomes, and the risks they pose to the World Bank. Ideally, the decision maker in each case should be the unit that benefits the most from a high-quality analytical product and suffers the most from a flop. This is the approach currently used by the World Bank for research products, which are mainly managed by DEC with an eye on its academic standing and without much interference from other units.

To some extent, the current Accountability and Decision-Making (ADM) framework does the same for other analytical tasks, spelling out clear protocols on how the requesting unit and the responsible unit should interact over critical steps in the process. But the ADM’s focus is more on processes rather on incentives. A second practical recommendation of this paper would thus be to use its proposed tiering of analytical tasks to better align responsibilities in the case of high-return-high-risk tasks, and to streamline processes in the other cases (table 5).

Table 5. Allocating management responsibilities to units with the highest stakes

Tier	Main risk	Process	Decides	Concurs
Flagships	World Bank's reputation	Yes (ADM)	Managing Director	Chief Economist or delegate
Debate			Vice-President	
Extended core	Client relationship		Country Director	Regional Chief Economist
Other country				
Research	Technical quality	No (Agile)	Academia	No (Agile)
Capacity	Resource waste		Budget holder	
Internal				

Finally, a third recommendation is to conduct more regular assessments of the quality and impact of analytical work across the entire institution, along the lines of what was done in this paper. Without aiming to provide a judgment on any individual knowledge product, data-driven studies of this sort allow assessing the overall performance distribution, and to identify the correlates of success. They would also allow to distinguish performance across various dimensions, such as technical quality, external visibility, and perceptions by end-users. This information, in turn, could be used to periodically adjust and upgrade knowledge management across the entire institution.

References

- Alfaro, Laura, Alan Auerbach, Mauricio Cardenas, Takatoshi Ito, Sebnem Kalemli-Ozcan and Justin Sandefur. 2021. "Doing Business: External Panel Review – Final Report". *Unpublished manuscript*. Washington, DC: The World Bank.
- Banerjee, Abhijit, Angus Deaton, Nora Lustig, Ken Rogoff, and Edward Hsu. 2006. "An Evaluation of World Bank Research, 1998-2005". *Unpublished manuscript*. Washington, DC: The World Bank.
- Dethier, Jean-Jacques. 2009. "World Bank Policy Research: A Historical Overview". *Policy Research Working Paper* 5000. Washington, DC: The World Bank.
- Doemeland, Doerte, and James, Trevino. 2014. "Which World Bank Reports are Widely Read?" *Policy Research Working Paper* 6851. Washington, DC: The World Bank Group.
- Knack, Stephen, Bradley C. Parks, Ani Harutyunyan, and Matthew DiLorenzo, Matthew. 2020. "How Does the World Bank Influence the Development Policy Priorities of Low-Income and Lower-Middle-Income Countries?" *Policy Research Working Paper* 9225. Washington, DC: The World Bank.
- Manuel, Trevor, Carlos Arruda, Jihad Azour, Chong-en Bai, Timothy Besley, Dong-Sung Cho, Sergei Guriev, Huguette Labelle, Jean Pierre Landau, Arun Maira and Hendrik Wolff. 2013. "Independent Panel Review of the Doing Business Report". *Unpublished manuscript*. Washington, DC: The World Bank.
- Ravallion, Martin, and Adam Wagstaff. 2011. "On Measuring Scholarly Influence by Citations." *Scientometrics* 88(1): 321-337.
- RePEc. 2022. "Top 10% Institutions and Economists in the Field of Development, as of August 2022". *Research Papers in Economics*. Retrieved on September 22, 2022, from <https://ideas.repec.org/top/top.dev.html>
- United Nations, Department of Economic and Social Affairs, Population Division. 2020. World Population Prospects 2020, Online Edition.
- World Bank. 2014. "World Bank Group Implements Major Reorganization". *World Bank Group Archives*. July 1, 2014. Retrieved on September 21, 2022, from <https://timeline.worldbank.org/event/3371>.
- World Bank. 2018. "Advisory Services and Analytics (ASA): Changing Patterns, Trends and Issues for Discussion". *Powerpoint presentation*. Washington, DC: The World Bank.
- World Bank. 2019. "Moving Toward a More Strategic and Impactful ASA Agenda". *Powerpoint presentation*. Washington, DC: The World Bank.

Appendix

Table A.1. Document types included in the consolidated database

Accounting and Auditing Assessment (ROSC)	6
Annual Report	3
Brief	1179
CAS Document	47
CAS Progress Report	3
Country Economic Memorandum	5
Country Engagement Note	10
Country Environmental Analysis (CEA)	2
Country Partnership Framework	23
Economic Updates and Modeling	10
ESMAP Paper	71
Financial Sector Assessment Program (FSAP)	16
IEG Evaluation	2
Information Notice	1
Insolvency Assessment (ROSC)	1
Investment Climate Assessment (ICA)	2
Journal Article	349
Manual	1
Newsletter	5
Other	11780
Other Agricultural Study	1
Other Education Study	3
Other Financial Sector Study	1
Other Poverty Study	4
Other Public Sector Study	1
Other Social Protection Study	2
Policy Note	18
Policy Research Working Paper	1445
Poverty Assessment	3
Procurement Plan	1
Program-for-Results Environmental and Social Systems Assessment	1
Program-for-Results Fiduciary Systems Assessment	4
Program-for-Results Technical Assessment	4
PSD, Privatization and Industrial Policy	1
Public Expenditure Review	15
Publication	346
Report	184
Systematic Country Diagnostic	75
Viewpoint	3
Working Paper	3340
Working Paper (Numbered Series)	270
World Development Indicators	1
World Development Report	36
Grand Total	19275

Table A.2. Determinants of ratings for all knowledge products

Explanatory variables	(1)	(2)	(3)	(4)	(5)	(6)
Total cost (K\$)	0.0170 *** (0.0050)	0.0161 *** (0.0047)	0.0144 *** (0.0034)	0.0167 *** (0.0042)	0.0160 *** (0.0041)	0.0145 *** (0.0031)
Trust funds (percent of cost)	-0.0275 *** (0.0060)	-0.0261 *** (0.0057)	-0.0350 *** (0.0075)	-0.0212 *** (0.0050)	-0.0203 *** (0.0049)	-0.0302 *** (0.0064)
Time to completion (months)	-0.0021 *** (0.0004)	-0.0019 *** (0.0004)	-0.0013 *** (0.0002)	-0.0017 *** (0.0003)	-0.0015 *** (0.0003)	-0.0010 *** (0.0002)
Top TTL		-0.1541 *** (0.0127)	-0.0421 *** (0.0128)		-0.1273 *** (0.0120)	-0.0375 *** (0.0128)
Good TTL		0.0604 *** (0.0165)	0.0618 *** (0.0161)		0.0537 *** (0.0164)	0.0566 *** (0.0162)
DEC			-0.0172 (0.0198)			0.0396 ** (0.0201)
Global Practice			0.2787 *** (0.0162)			0.2808 *** (0.0164)
Regional CE office			0.2601 *** (0.0413)			0.2069 *** (0.0407)
FY17	0.3090 *** (0.0104)	0.3055 *** (0.0103)	0.2959 *** (0.0101)	0.3026 *** (0.0101)	0.3000 *** (0.0100)	0.2930 *** (0.0098)
FY18	0.2571 *** (0.0107)	0.2553 *** (0.0103)	0.2557 *** (0.0092)	0.2645 *** (0.0097)	0.2628 *** (0.0095)	0.2618 *** (0.0088)
FY19	0.3070 *** (0.0088)	0.3096 *** (0.0086)	0.3121 *** (0.0082)	0.3073 *** (0.0085)	0.3094 *** (0.0083)	0.3120 *** (0.0080)
Region fixed effects	No	No	No	Yes	Yes	Yes
Flagships and debates (yes = 1)	-0.0098 (0.0088)	-0.0117 (0.0088)	-0.0368 *** (0.0090)	0.0661 *** (0.0092)	0.0617 *** (0.0092)	0.0351 *** (0.0095)
Country products (yes = 1)	-0.0008 *** (0.0001)	-0.0007 *** (0.0001)	-0.0006 *** (0.0001)	0.0001 ** (0.0001)	0.0001 ** (0.0001)	0.0001 ** (0.0001)
Constant	0.3831 *** (0.0109)	0.3833 *** (0.0104)	0.1099 *** (0.0178)	0.0360 * (0.0189)	0.0473 *** (0.0181)	-0.1902 *** (0.0230)
Observations	11,387	11,387	11,387	11,387	11,387	11,387
R-squared	0.176	0.188	0.233	0.221	0.229	0.266
F-test	312.7 ***	305.8 ***	413.1 ***	312.2 ***	306.7 ***	369.2 ***

Note: High, medium and low task ratings are computed as 1, 2/3 and 1/3 respectively. Robust standard errors are reported in parentheses. Statistically significant coefficients at the 1, 5 and 10 percent significance levels are indicated by three, two and one asterisks respectively.

Table A.3. Determinants of ratings for flagships and debates

Explanatory variables	(1)	(2)	(3)	(4)	(5)	(6)
Total cost (K\$)	0.0156 *** (0.0044)	0.0155 *** (0.0045)	0.0187 *** (0.0047)	0.0164 *** (0.0045)	0.0164 *** (0.0046)	0.0195 *** (0.0047)
Trust funds (percent of cost)	-0.0803 *** (0.0177)	-0.0799 *** (0.0177)	-0.0783 *** (0.0180)	-0.0669 *** (0.0183)	-0.0673 *** (0.0183)	-0.0681 *** (0.0185)
Time to completion (months)	-0.0022 *** (0.0003)	-0.0022 *** (0.0003)	-0.0022 *** (0.0003)	-0.0022 *** (0.0003)	-0.0022 *** (0.0003)	-0.0022 *** (0.0003)
Top TTL		-0.0501 * (0.0301)	-0.0451 (0.0305)		-0.0383 (0.0309)	-0.0298 (0.0312)
Good TTL		0.0668 * (0.0379)	0.0706 * (0.0372)		0.0529 (0.0379)	0.0600 * (0.0373)
DEC			-0.1850 *** (0.0670)			-0.1849 *** (0.0685)
Global Practice			-0.0121 (0.0474)			-0.0115 (0.0485)
Regional CE office			0.0196 (0.0641)			-0.0044 (0.0656)
FY17	0.3819 *** (0.0257)	0.3825 *** (0.0256)	0.3791 *** (0.0255)	0.3828 *** (0.0254)	0.3832 *** (0.0254)	0.3800 *** (0.0253)
FY18	0.3178 *** (0.0230)	0.3171 *** (0.0230)	0.3172 *** (0.0231)	0.3216 *** (0.0230)	0.3210 *** (0.0230)	0.3210 *** (0.0230)
FY19	0.4005 *** (0.0216)	0.4017 *** (0.0216)	0.4010 *** (0.0215)	0.4040 *** (0.0214)	0.4047 *** (0.0214)	0.4047 *** (0.0213)
Region fixed effects	No	No	No	Yes	Yes	Yes
Flagships (yes = 1)	-0.1622 *** (0.0334)	-0.1592 *** (0.0334)	-0.1395 *** (0.0336)	-0.1500 *** (0.0340)	-0.1477 *** (0.0340)	-0.1279 *** (0.0344)
Constant	0.2741 *** (0.0180)	0.2734 *** (0.0183)	0.2840 *** (0.0485)	0.2432 *** (0.0206)	0.2428 *** (0.0209)	0.2573 *** (0.0510)
Observations	2,077	2,077	2,077	2,077	2,077	2,077
R-squared	0.240	0.242	0.246	0.248	0.250	0.254
F-test	101.3 ***	80.9 ***	62.8 ***	63.6 ***	56.0 ***	47.9 ***

Note: High, medium and low task ratings are computed as 1, 2/3 and 1/3 respectively. Robust standard errors are reported in parentheses. Statistically significant coefficients at the 1, 5 and 10 percent significance levels are indicated by three, two and one asterisks respectively.

Table A.4. Determinants of ratings for country products

Explanatory variables	(1)	(2)	(3)	(4)	(5)	(6)
Total cost (K\$)	0.0027 (0.0062)	0.0025 (0.0062)	0.0009 (0.0063)	0.0046 (0.0063)	0.0043 (0.0064)	0.0029 (0.0063)
Trust funds (percent of cost)	-0.0209 *** (0.0055)	-0.0208 *** (0.0055)	-0.0324 *** (0.0089)	-0.0180 *** (0.0052)	-0.0178 *** (0.0052)	-0.0296 *** (0.0078)
Time to completion (months)	0.0000 (0.0003)	0.0001 (0.0003)	0.0002 (0.0004)	0.0001 (0.0003)	0.0002 (0.0004)	0.0003 (0.0004)
Top TTL		-0.0836 *** (0.0241)	-0.0658 *** (0.0246)		-0.0866 *** (0.0241)	-0.0682 *** (0.0246)
Good TTL		0.0397 * (0.0205)	0.0327 (0.0202)		0.0403 ** (0.0206)	0.0336 * (0.0202)
DEC			0.1981 *** (0.0541)			0.1928 *** (0.0548)
Global Practice			0.3912 *** (0.0186)			0.3922 *** (0.0186)
Regional CE office			0.4991 *** (0.1361)			0.4900 *** (0.131x)
FY17	0.2900 *** (0.0122)	0.2892 *** (0.0122)	0.2908 *** (0.0118)	0.2881 *** (0.0123)	0.2872 *** (0.0123)	0.2888 *** (0.0118)
FY18	0.2623 *** (0.0118)	0.2620 *** (0.0118)	0.2679 *** (0.0111)	0.2627 *** (0.0118)	0.2625 *** (0.0118)	0.2685 *** (0.0111)
FY19	0.2909 *** (0.0106)	0.2912 *** (0.0105)	0.2964 *** (0.0100)	0.2933 *** (0.0105)	0.2937 *** (0.0105)	0.2987 *** (0.0099)
Region fixed effects	No	No	No	Yes	Yes	Yes
Extended core (yes = 1)	0.0581 *** (0.0113)	0.0565 *** (0.0113)	0.0655 *** (0.0110)	0.0488 *** (0.0117)	0.0475 *** (0.0117)	0.0580 *** (0.0112)
Constant	0.3063 *** (0.0097)	0.3072 *** (0.0100)	-0.0629 *** (0.0206)	0.2229 (0.2457)	0.2239 (0.2458)	-0.0809 (0.2029)
Observations	6,784	6,784	6,784	6,784	6,784	6,784
R-squared	0.142	0.144	0.199	0.186	0.189	0.244
F-test	162.7 ***	130.7 ***	147.1 ***	93.6 ***	83.9 ***	105.4 ***

Note: High, medium and low task ratings are computed as 1, 2/3 and 1/3 respectively. Robust standard errors are reported in parentheses. Statistically significant coefficients at the 1, 5 and 10 percent significance levels are indicated by three, two and one asterisks respectively.

Table A.5. Determinants of downloads for all knowledge products

Explanatory variables	(1)	(2)	(3)	(4)	(5)	(6)
Total cost (K\$)	1.725 ** (0.761)	1.795 ** (0.767)	1.892 ** (0.765)	1.829 ** (0.760)	1.838 ** (0.762)	1.976 *** (0.762)
Trust funds (percent of cost)	-7.259 * (4.390)	-7.190 * (4.246)	-3.112 (2.027)	-8.514 * (5.062)	-8.436 * (4.939)	-4.343 * (2.573)
Time to completion (months)	-0.034 (0.026)	-0.059 * (0.035)	-0.065 * (0.037)	-0.032 (0.026)	-0.053 * (0.029)	-0.076 ** (0.037)
Top TTL		26.68 ** (11.53)	16.24 (13.25)		25.39 ** (11.47)	13.57 (13.41)
Good TTL		53.63 (46.44)	45.89 (42.11)		51.59 (45.43)	44.19 (41.30)
DEC			-2.486 (9.396)			-0.980 (8.619)
Global Practice			-5.003 ** (2.438)			-5.227 ** (2.577)
Regional CE office			322.9 * (172.8)			321.0 * (172.3)
FY17	9.053 *** (3.417)	8.767 ** (3.520)	7.115 ** (3.626)	9.771 *** (3.503)	9.449 *** (3.535)	7.920 ** (3.562)
FY18	5.988 *** (2.102)	5.493 ** (2.194)	4.694 ** (2.372)	6.922 *** (2.185)	6.444 *** (2.215)	5.139 ** (2.374)
FY19	20.80 * (11.75)	20.41 * (11.70)	18.85 * (10.70)	19.95 * (11.32)	19.64 * (11.33)	18.04 * (10.33)
Region fixed effects	No	No	No	Yes	Yes	Yes
Flagships and debates (yes = 1)	34.22 ** (14.16)	34.49 ** (14.13)	19.54 *** (6.757)	38.38 ** (15.19)	38.93 *** (15.09)	21.10 *** (6.181)
Country products (yes = 1)	-0.064 (0.056)	-0.067 (0.054)	-0.027 (0.038)	-0.010 (0.055)	-0.013 (0.057)	-0.009 (0.054)
Constant	6.839 (5.903)	3.731 (3.973)	2.370 (3.978)	-11.19 (9.667)	-14.83 (10.33)	-2.187 (10.49)
Observations	11,387	11,387	11,387	11,387	11,387	11,387
R-squared	0.004	0.007	0.028	0.007	0.009	0.030
F-test	3.14 **	2.86 **	2.65 **	2.15 **	2.21 **	1.87 **

Note: Downloads are measured per million people in the population of the geographic area of the intended audience. Robust standard errors are reported in parentheses. Statistically significant coefficients at the 1, 5 and 10 percent significance levels are indicated by three, two and one asterisks respectively.

Table A.6. Determinants of downloads for flagships and debates

Explanatory variables	(1)	(2)	(3)	(4)	(5)	(6)
Total cost (K\$)	4.627 (3.639)	3.812 (3.904)	2.819 (3.402)	7.051 ** (3.552)	6.154 * (3.579)	4.128 (3.314)
Trust funds (percent of cost)	-34.14 * (18.22)	-28.19 * (15.50)	6.475 (11.76)	-41.57 * (21.45)	-37.12 * (19.96)	-8.537 (11.02)
Time to completion (months)	-1.226 ** (0.552)	-1.179 ** (0.474)	-1.002 ** (0.400)	-1.287 ** (0.575)	-1.230 ** (0.497)	-1.041 ** (0.420)
Top TTL		126.1 ** (62.27)	81.63 (58.92)		71.65 (65.04)	34.90 (69.85)
Good TTL		290.8 (253.1)	247.4 (233.3)		277.4 (243.6)	242.0 (228.6)
DEC			-18.28 (105.0)			11.09 (93.45)
Global Practice			-11.48 (13.94)			-1.556 (15.84)
Regional CE office			386.1 * (199.7)			365.7 * (187.8)
FY17	18.22 (13.30)	5.255 (15.80)	-1.474 (18.77)	23.01 * (13.31)	13.29 (14.20)	5.539 (16.39)
FY18	3.460 (17.68)	0.301 (19.96)	-0.870 (20.22)	8.670 (15.45)	5.442 (17.36)	2.512 (18.33)
FY19	120.5 * (67.32)	117.8 * (66.41)	106.0 * (59.68)	115.1 * (63.63)	113.9 * (63.56)	102.6 * (57.50)
Region fixed effects	No	No	No	Yes	Yes	Yes
Flagships (yes = 1)	355.3 * (192.1)	347.9 * (190.1)	324.1 * (185.4)	368.7 * (195.6)	364.2 * (195.3)	329.6 * (185.8)
Constant	42.40 ** (19.14)	20.06 * (10.97)	-1.981 (14.63)	15.53 (12.98)	-0.298 (14.85)	-11.33 (21.03)
Observations	2,077	2,077	2,077	2,077	2,077	2,077
R-squared	0.032	0.046	0.067	0.054	0.065	0.082
F-test	1.84	2.01	1.63	1.05*	1.26	1.18

Note: Downloads are measured per million people in the population of the geographic area of the intended audience. Robust standard errors are reported in parentheses. Statistically significant coefficients at the 1, 5 and 10 percent significance levels are indicated by three, two and one asterisks respectively.

Table A.7. Determinants of downloads for country products

Explanatory variables	(1)	(2)	(3)	(4)	(5)	(6)
Total cost (K\$)	0.400 (0.342)	0.401 (0.342)	0.419 (0.346)	0.334 (0.320)	0.331 (0.320)	0.344 (0.324)
Trust funds (percent of cost)	-0.357 (0.692)	-0.354 (0.692)	-0.112 (0.674)	-0.479 (0.698)	-0.478 (0.699)	-0.238 (0.664)
Time to completion (months)	0.0036 (0.0109)	0.0036 (0.0109)	0.0021 (0.0109)	0.0003 (0.0098)	0.0002 (0.0098)	-0.0010 (0.0101)
Top TTL		0.418 (1.417)	0.369 (1.382)		0.486 (1.429)	0.467 (1.383)
Good TTL		2.449 (2.446)	2.584 (2.447)		2.633 (2.434)	2.763 (2.435)
DEC			-7.118 * (3.777)			-7.193 * (3.796)
Global Practice			-7.940 ** (3.083)			-7.755 ** (3.064)
Regional CE office			-9.660 *** (3.145)			-9.199 *** (3.204)
FY17	3.795 (2.835)	3.770 (2.838)	3.729 (2.839)	3.869 (2.842)	3.843 (2.845)	3.802 (2.846)
FY18	-0.205 (0.700)	-0.237 (0.685)	-0.339 (0.703)	-0.182 (0.698)	-0.216 (0.683)	-0.315 (0.701)
FY19	-0.833 * (0.471)	-0.827 * (0.471)	-0.907 * (0.489)	-0.891 * (0.485)	-0.882 * (0.485)	-0.954 * (0.501)
Region fixed effects	No	No	No	Yes	Yes	Yes
Extended core (yes = 1)	0.341 (0.937)	0.331 (0.942)	0.185 (0.893)	1.004 (1.010)	0.987 (1.015)	0.815 (0.951)
Constant	1.835 *** (0.588)	1.721 *** (0.551)	9.220 *** (3.166)	-1.673 (1.198)	-1.650 (1.219)	4.321 * (3.419)
Observations	6,784	6,784	6,784	6,784	6,784	6,784
R-squared	0.001	0.002	0.003	0.002	0.003	0.004
F-test	4.04 ***	3.31 ***	3.30 ***	2.79 ***	2.51 **	2.48 ***

Note: Downloads are measured per million people in the population of the geographic area of the intended audience. Robust standard errors are reported in parentheses. Statistically significant coefficients at the 1, 5 and 10 percent significance levels are indicated by three, two and one asterisks respectively.