

An Evaluation of World Bank Research, 1998 - 2005

Abhijit Banerjee (MIT)
Angus Deaton (Princeton), Chair
Nora Lustig (UNDP)
Ken Rogoff (Harvard)

with Edward Hsu (IFC)

and assisted by

Daron Acemoglu (MIT), Joshua Angrist (MIT), Marianne Bertrand (U. Chicago), Timothy Besley (LSE), Nancy Birdsall (Center for Global Development), Francesco Caselli (LSE), Peter Diamond (MIT), Esther Duflo (MIT), Sebastian Edwards (UCLA), Marcel Fafchamps (Oxford), Andrew Foster (Brown), Sebastian Galiani (Washington U), Geoffrey Heal (Columbia), Edward Glaeser (Harvard), Michael Kremer (Harvard), Murray Leibbrandt (U Cape Town), Justin Lin (China Center for Economic Research), Jonathan Morduch (NYU), Nina Pavcnik (Dartmouth), Gordon Hanson (UCSD), Antoinette Schoar (MIT), Jan Svejnar (Michigan), Christopher Udry (Yale), and Martin Wittenberg (U. Cape Town)

September 24, 2006

We are grateful to the staff of the Research Support Group of the World Bank, particularly Jean-Jacques Dethier, Clara Else, Shiva Makki, Anupa Bhaumik, Fadima Savadogo, Trinidad Angeles, Evelyn Alfaro-Bloch, and Thi Trang Linh Phu for extensive support and assistance. The views expressed in this document are those of the panel members acting in a personal capacity and are not meant to represent the views of the institutions with which the members of the panel are affiliated.

TABLE OF CONTENTS

Executive Summary	5
Chapter 1. The World Bank and Research	11
Why should the World Bank do Research?	11
<i>The World Bank’s Advantage in Research</i>	14
Problems of doing research in the World Bank	18
Implications for the evaluation	22
Chapter 2. How and where research is done in the World Bank	24
How Research is Managed and Organized	24
<i>Development Economics Vice Presidency (DEC)</i>	27
<i>Regions</i>	28
<i>Networks</i>	28
<i>World Bank Institute</i>	29
Types of Products	29
<i>Journal Articles</i>	30
<i>Policy Research Working Papers</i>	30
<i>Analytical Tools</i>	31
<i>Policy Research Reports</i>	31
<i>Data Products</i>	32
<i>Annual Bank Conference on Development Economics</i>	32
<i>Special Flagship Reports</i>	32
<i>World Development Reports</i>	33
The Bank’s Research Portfolio	33
<i>Finance</i>	34
<i>Growth and Investment</i>	35
<i>Human Development and Public Services</i>	35
<i>Infrastructure and Environment</i>	35
<i>Poverty</i>	36
<i>Rural Development</i>	36
<i>Trade and International Integration</i>	36
Chapter 3. Assessing the Quality of World Bank Research	37
How the evaluation was organized	40
<i>A note on citation analyses</i>	44
Evaluation results: an overview	46
Highlights: Bank research at its best	46
Important topics, but with serious shortcomings in execution and conclusions ...	51
<i>Globalization, aid, and poverty</i>	52
<i>Pensions and insurance</i>	58
<i>Infrastructure</i>	59
<i>Poverty mapping</i>	59
<i>Civil war</i>	64
<i>Finance and growth</i>	64
General themes: strengths and weaknesses	66
<i>Execution and methods</i>	66

<i>Computable general equilibrium techniques</i>	67
<i>Project evaluation</i>	68
<i>Analytical narratives</i>	69
<i>Use of non-Bank consultants and researchers</i>	70
<i>Heterogeneity of quality, jumping to conclusions, and self-citation</i>	71
<i>Missing areas?</i>	73
<i>Academic versus policy agendas</i>	74
<i>Dissemination: closing the loop</i>	75
The World Development Reports	75
<i>A disclaimer</i>	75
<i>Background</i>	76
<i>Many strengths</i>	78
<i>Some weaknesses</i>	81
Chapter 3: Annex 1: Further remarks on poverty mapping	86
Chapter 3: Annex 2: Analysis of evaluators' scores	89
Table 1: Averages scores by aspects of research	97
Table 2: DEC v non-DEC and flagships versus non-flagships by aspects of research	98
Table 3. Strengths and weaknesses of research	99
Chapter 4. Evaluator comments by area	101
<i>Macroeconomics and growth</i>	101
<i>Fiscal policy, public sector management, and governance</i>	103
<i>Trade and international economics</i>	103
<i>Poverty and social welfare</i>	105
<i>Human development (health, education, population, employment)</i>	108
<i>Finance and private sector development</i>	110
<i>Agriculture and rural development</i>	113
<i>Infrastructure and urban development</i>	115
<i>Environment</i>	116
<i>Flagship reviews: pensions and insurance, Doing Business, and transition</i>	117
Chapter 5. What we learned from the interviews	119
<i>Introduction and Overview</i>	119
<i>The view from inside: how World Bank research perceived by Bank researchers? ..</i>	122
<i>Looking back: The views of some past leaders of the research department</i>	126
<i>The view from operations</i>	130
<i>The view from outside: what did we hear from policy people and senior academics in borrowing countries?</i>	135
Chapter 6. World Bank research: exploring institutional options	137
Problem areas	137
<i>Budget squeeze</i>	137
<i>The Bank should be able to produce a lower proportion of research that is neither policy relevant nor academically distinguished</i>	139
<i>The fundamental tension between the Bank's role as an advocate of good policies and a producer of new policy ideas</i>	140
<i>The balance between rigor and relevance</i>	141
<i>Balance between responsiveness and independence</i>	142

<i>Data collection and maintenance</i>	143
<i>Statistical and econometric expertise</i>	144
<i>Too Many Thick Volume Flagship Reports</i>	145
<i>More Support for Institution Based Research in Developing Countries</i>	147
Addressing the problem areas	148
<i>An overarching recommendation: Learning what works and telling the world</i>	148
<i>Financing Research and protecting its independence and objectivity</i>	148
<i>Control mechanisms for more consistent pruning of weak research</i>	150
<i>Improving Flagship Reports</i>	152
<i>Strengthening interactions with academics and bringing in new ideas</i>	153
<i>Dealing with the Bank’s overly diffuse structure for allocating and planning research</i>	154
<i>Making data truly a public good</i>	157
<i>Improved cost accounting for research</i>	159
<i>Creating a More Formal Mechanism for Research Replication</i>	160
Summary of recommendations	161

Executive Summary

This evaluation of World Bank research between 1998 and 2005 was carried out by a panel consisting of Abhijit Banerjee (MIT), Angus Deaton (Princeton, chair), Nora Lustig (UNDP), and Kenneth Rogoff (Harvard.) The panel selected a large random sample of research projects, which were read and assessed by a team of 25 evaluators. Panel members also solicited views from current and past Bank staff, as well as from policy makers and academics in developing countries.

Based on the evidence we assembled, the interviews we conducted, and our own consideration, the panel concluded that the World Bank needs a research department, and that its research needs cannot be fully met by hiring in from the outside. Research is a central part of quality control in the Bank, and is crucial to its claim to be a “Knowledge Bank.” Without a research-based ability to learn from its projects and policies, the Bank could not maintain its role as the world’s leading development agency. The 2.5 percent of its administrative budget that the Bank spends on research is surely too low given the multiplicity of tasks that research is expected to fulfill, including the generation of new knowledge about development, the collection and dissemination of data, the generation of knowledge to support guide Bank strategy, operational support, and capacity building in client countries. As the world becomes richer, and already today among middle income countries, the need for high-quality, research-based advice will only become stronger as the need for Bank lending diminishes.

The multiple tasks of Bank research are not always consistent with one another, and we believe that the Bank’s Chief Economists and their research staffs deserve considerable credit for the way that they have fulfilled their obligations over the past seven years. They have done so in a period when new hiring has been severely limited, and where the salaries of Bank economists have fallen rapidly relative to those in academia. Bank researchers have produced innovative and important new research that has maintained the Bank’s position as the intellectual leader among development agencies. At the same time they have provided extensive support to their colleague in operations; indeed researchers in the Bank’s research department devote 30 percent of their time to such operational “cross-support.”

Bank researchers and their consultants produced nearly 4,000 papers, books, and reports between 1998 and 2005. Bank researchers regularly publish in the leading academic journals in economics, and more extensively in the leading field journals in development. The Development Economics group (DEC) is also responsible for the annual *World Development Report*, which is widely read by the development community, and which has sometimes had a major effect on development thinking. The Bank also publishes a large numbers of policy documents and reports that summarize the state of the art in various policy areas and that are designed to communicate and disseminate research to policymakers and their advisors. Research is done throughout the Bank, by economists working in the regions and in the Bank’s networks, as well as, most importantly, in the research group of DEC.

Our evaluators and the panel found some outstanding work in the Bank's portfolio. Bank economists have led the world in the measurement of poverty and inequality, including inequality in health. Pioneering research on the organization of and delivery of educational and health services is changing the way we think about these issues and the way that the Bank lends money for such projects. There is important work on monitoring the environment. The Bank has been a world leader in the collection of new data, including the long-established Living Standards Measurement Surveys, the joint household survey project with the Inter American Development Bank called MECOVI, as well as the more recent Business Environment and Economic Performance surveys in the transition countries, and the Investment Climate and Doing Business surveys. The Bank's data group collates the World Development Indicators, which is the most important single database for development research, and it has recently taken on board the International Comparison Project, which is central for the measurement of economic growth, for poverty, and comparative measures of development around the world.

Bank researchers have also done extremely visible work on globalization, on aid effectiveness, and on growth and poverty. In many ways they have been the leaders on these issues. But the panel had substantial criticisms of the way that this research was used to proselytize on behalf of Bank policy, often without taking a balanced view of the evidence, and without expressing appropriate skepticism. Internal research that was favorable to Bank positions was given great prominence, and unfavorable research ignored. There were similar criticisms of the Bank's work on pensions, which produced a great deal that was useful, but where balance was lost in favor of advocacy. In these cases, we believe that there was a serious failure of the checks and balances that should separate advocacy and research. The panel endorses the right of the Bank to strongly defend and advocate its own policies. But when the Bank leadership selectively appeals to relatively new and untested research as hard evidence that its preferred policies work, it lends unwarranted confidence to the Banks' prescriptions. Placing fragile selected new research results on a pedestal invites later recrimination that undermines the credibility and usefulness of all Bank research.

Data collection and dissemination is another area where the Bank has many great achievements but there are also many problems. The panel sees the Bank's data work as central to its mission of learning from development. It is not only the basis of most Bank research, but it automatically scales up Bank work by permitting research by others, an increasingly large number of whom are in developing countries. Yet data activities are organized haphazardly, whether in collection, archiving, or dissemination. The Development Economics data group is not as centrally involved with researchers in the collection and dissemination of Bank data as is desirable. The Bank website is often of poor quality and difficult to use, not only for accessing data, but even for the relevant publications and reports. The Bank has no coherent policy for data release, for its own researchers, nor for client countries to which it provides support in data collection. Too little has been done to build on the early success of the Living Standards Measurement Surveys to help build internationally comparable data on such central topics as poverty or mortality. Without improvements here, there is a long term threat to the Bank's (and the world's) ability to monitor the income and health dimensions of world poverty.

Bank research has become predominantly empirical, with routine use of econometric and statistical methods. This is as it should be; learning from experience requires statistical analysis. Yet the panel, while recognizing that there has already been substantial movement in the right direction, believes that Bank could still make more use of randomized experiments in those cases where they are possible, for example, for many projects in the social sector. With or without randomized trials, Bank researchers are not often enough involved in the early stages of project planning, where they can be instrumental in laying the foundations for successful learning after completion. Without such efforts, the Bank cannot routinely learn from its own experience. We welcome the initiatives in these areas that are underway in the Bank, but press the need for more. The problem of keeping abreast of new approaches applies to a broad range of applications, not just new uses of randomized experiments. We suspect that management has not always kept ahead of researchers in their understanding and familiarity with statistical and econometric methods, and that this has sometimes contributed to the failure to appropriately interpret and manage research results. The Bank's misplaced confidence in cross country regressions on growth, poverty and aid, is a case in point. Another is its lack of a full understanding of the limitations of the innovative methods developed by Bank researchers to estimate poverty for small areas; once again, results were sold without appropriate caution and qualification. Although the quality of statistical work is a Bank wide issue, DEC (and within it, perhaps the DEC data group) is the obvious home for statistical and econometric leadership. The Bank needs a "central statistical office" and should consider whether it needs a chief statistician to head it.

Our evaluators generally found that Bank research was well-targeted towards important topics, but was often weak on execution and technique. While it is desirable for Bank technique to be behind the frontier, there has often been too large a gap. Some technically-flawed projects have run for years, and have been incorporated into country work without appropriate certification and review. The evaluators repeatedly found that too large a fraction of Bank research was undistinguished, in the sense that it had neither great relevance to policy nor claim to academic distinction. These are subjective judgments, but our evaluators are distinguished development economists, and their views were very similar to one another. Their judgments did not refer to the lack of good papers in good journals, many of which were innovative and important by any standard. Nor were any of them counting citations. The concern was with the large fraction of papers that, on reading, did not seem to be very useful from the perspective of either an academic or a policymaker.

Bank researchers in the Development Economics Group (DEC) are expected to publish two academic papers a year, and this mechanism helps guarantee quality and protect the Bank's intellectual standing. But the cost, at least within DECRG, is a large number of less than outstanding papers driven too much by the concerns of journals and their referees and too little by the policy needs of the Bank. Nor do these papers make use of the Bank's comparative advantages of local knowledge and a constant stream of important new problems. At the same time, there is great pressure for researchers to demonstrate policy relevance, which frequently leads to drawing conclusions that are not

supported by the evidence. There is too much self-citation. Some of the very best and very worst work was done jointly with outside consultants whose quality was clear in advance. The evaluators generally gave higher scores to research in DEC than to research done elsewhere in the Bank, although they scored the non-DEC flagships as highly as they did regular DEC research.

The *World Development Reports* have sometimes been instrumental in changing the way that the world thinks about some aspect of development, such as poverty, health, or population. In recent years, they have, to an extent, become the victims of their own success. Because they are seen as so important, they must incorporate the views of large numbers of people, inside and outside the Bank. In consequence, they often seek to minimize conflict and to emphasize “win-win” situations instead of trade-offs. They often lack sharpness and focus, and are sometimes incoherent, especially when it proves impossible to reconcile the views of the various commentators and authors. They are also extraordinarily expensive, absorbing about ten percent of the resources of the research department. Even so, the panel thinks they should probably continue. They provide the Chief Economist with a highly visible vehicle for summarizing and disseminating research on issues that he or she considers to be important, and their regular appearance contributes to the Bank’s standing in the development community even if, to some extent, they are trading on their past reputation.

The panel gave considerable thought to what should be expected of Bank researchers in terms of academic publication. Satisfying the requirements of academic editors and their reviewers is not the main business of the Bank. But without an expectation of publication, the Bank could not maintain its reputation as the leading thinker in economic development. Nor would it be able to attract the high quality researchers that it needs to think about and to help address the many problems of development. Yet too much pressure to publish leads researchers to ignore important policy issues in favor of an academic style that is sometimes of limited value. We believe that the tension here is a fundamental one that will always be faced by the research managers in the Bank. The “two publications a year” norm seems to us to be a reasonable mechanism, as is the requirement that researchers in DEC spend 30 percent of their time in operational support. We also recognize that the publication rule will lead to a substantial body of work of the kind noted above, that is successful neither academically, nor in policy relevance. This is perhaps the inevitable cost of an imperfect quality-control mechanism.

Even so, we believe that there has been too much of this sort of work over the review period. Bank research has not been monitored and evaluated as often as is desirable. The fact that our evaluation is the first in seven years is not unrelated to some of the problems that we have found. More regular evaluations would permit early termination of bad projects, and would help limit the long tail of undistinguished work. The Bank needs better tracking systems to link research expenditures to research outputs; currently it is not even tracking outputs so that it is impossible to know exactly what has been produced. The Bank needs to encourage better links with academics, both in the selection of outside researchers as consultants, which is currently too haphazard and decentralized, and in fostering regular interchanges through visitors and conferences around key topics.

Researchers should not be hiring consultants whose track records gives clear advance indication that they are unlikely to produce good work; that they do so suggests a failure of monitoring and management.

While we do not think it makes sense for the Bank to contract out all or even most of its research, for example by issuing requests for proposals, we think that it should consider using this mechanism on occasion where Bank expertise is not available. We noted how little of the research that we saw involves joint work with researchers from developing countries. While we are acutely aware of the difficulties of doing better, we emphasize the importance of attempting to do so, perhaps through greater institutional support, or by supporting highly trained immigrant economists in the US and Europe to spend time in their native countries.

We are also concerned with quality control over the Bank's large number of "flagship" publications, here taken to be the *World Development Reports* and the DEC and non-DEC major topic studies to which the term is applied. These reports are sometimes enormously influential (though we suspect that many just gather dust) and they are the vehicles where the line between the Banks' advocacy role and its role in producing new research ideas becomes particularly blurred. The large number of flagship reports makes it virtually impossible for management to exert sufficient quality control precisely where it is most needed. The Chief Economist's office, even if it were vested with sign-off authority on all flagships, lacks the time and resources to adequately vet them. We believe that the Bank produces too many of these reports. It should find a mechanism for better quality control of a smaller number, either by extending the Chief Economist's authority, and giving him or her resources to undertake the quality control, or by requiring some sort of outside review, or both.

In spite of the centrality of the research to the Bank's mission, it is continually necessary to lobby for research, and to protect basic research on development issues, especially where the payoffs are not immediate. The panel believes that there would be great benefits to endowing the Bank's development policy research, which could be done using a small fraction of the Bank's cumulated retained earnings. Without such insulation, there is a risk that it will degenerate into pure advocacy of the type that has become all too prevalent in the global poverty debate. The Bank must maintain its distinction in research.

Chapter 1. The World Bank and Research

Why should the World Bank do Research?

The World Bank is one of the most important centers of research in development economics today. It spends approximately two and a half percent of its total budget on research and its research department, including 93 researchers and more than 30 support staff, is by far the biggest single group of high-quality researchers in development economics. There are also prominent researchers outside the research group, including some who are located in the country offices. Many of these researchers are world leaders in development research, and some of the most important new thinking in development has come from World Bank researchers. The World Bank Chief Economists have been among the world's leading scholars, including a Nobel Laureate. The Bank has also been the single most important producer and collator of data about economic development, and Bank data support a vast amount of research inside the Bank, as well as by researchers, policy analysts and governments around the world.

Why does the Bank need to be such a major player in development economics? A superficial answer is that all commercial banks have a research department, which helps them figure out what they should finance and what they ought to stay away from. Since the World Bank finances development, it needs development research. This analogy, while sometimes used to justify Bank research is a poor one. Commercial banks care about the success of the projects they fund because they want to be in a position to collect. By contrast, repayment to the World Bank is never directly tied to the success of any specific initiative. Its loans are typically guaranteed by the general revenues of the government. What the World Bank does care very much about is helping poor and

middle income countries find ways to achieve rapid sustainable growth while achieving significant reductions in poverty. Its interests are not really those of a lender so much as that of a partner in development. And here, the ability of researchers to enhance the Bank's vision of global best development practices, is invaluable.

Thus a far more compelling argument starts from the Bank's role as the world's premier development agency. The Bank exists ultimately to promote development, which requires a base of knowledge, much of which must be generated by the Bank itself. As is well known, there is an on-going effort to reposition the World Bank as the "Knowledge Bank," with lending operations playing a reduced role, and the Bank playing a more important role as a source of policy knowledge. In many ways this is responding to the changing demand for the Bank's services. We already see that a number of middle income countries like Mexico, or even countries approaching middle income, like India, either do not really need the Bank as a lender or are moving in that direction. On the other hand, a number of policy makers from these countries told us that they really value the Bank as a source of high-quality technical advice on complex issues. This shift in the demand for the Bank's services will continue as the world gets richer, and many people within the Bank see a more central role for research and advice in the Bank's future. These arguments apply beyond research; countries often value the Bank's expertise and experience in such matters as procurement, international bidding, or project level financial systems.

Moreover, even in the case of the poorest countries, where access to IDA loans and other credit from the Bank remains economically important, there is now an on-going discussion of whether the Bank ought to move to a model where it is less a lender and

more a helping hand, dispensing grants and advice. How the Bank will finance its essential research functions in this evolving environment is an important challenge we will take up towards the end of this report. Suffice for now to say that we regard the financing challenge as secondary to the broader question of how to maintain and strengthen the Bank's research output, which is our main concern here.

Regardless of how it finances knowledge creation, the Bank could in principle devote its research resources to funding researchers in universities and other organizations to work on topics that are important to its mission. However, there are a number of compelling reasons why a large part of this research needs to be done within the Bank.

The most obvious reason is that the Bank needs to be able to seamlessly integrate good economics and as well as research ideas into its day to day activities. Bank staff face constant questions and decisions on how to fine tune their development policy advice, and how to best expend the Banks' resources. To this end, there is no substitute for a strong research group that is deeply integrated into the Bank's decision making. In addition, it is the backing of good economists that lends credibility (and substance) to the advice it gives to countries.

The Bank's economists also play a central role in addressing the broad development community, within which the World Bank has always been seen as an intellectual leader. In part, this is because the Bank is the largest funder of development, and is involved in the whole range of development policies and projects, so that it has the most experience from which to learn. One of the tasks of Bank economists is to learn from that experience and to communicate their findings broadly, to other developing countries, who want to learn from others' failures and successes, and to the community of funders and scholars

who need to understand what works and what does not work. More broadly still, Bank economists are in a strong position to think about the “big” questions, such issues as how to reduce poverty, how to help Africa grow faster, how to balance social sectors like health and education with more narrowly economic investments, or whether and under what circumstances aid works. There is a long tradition of such big thinking from the Bank, which has sometimes been hugely influential on global ideas about development. Indeed, Bank researchers almost certainly have more influence on Bank operations *indirectly*, through their influence on the broad community, as *directly*, through their advice on particular programs and projects.

The World Bank’s Advantage in Research

There are many areas in which development research can either only be done by the Bank, or where Bank researchers have a substantial advantage. Among these, the most important are:

1. The Bank has an ability to collect data in collaboration with the statistical agencies in member countries. Bank researchers often have access to data that could not readily be granted to independent or commercial researchers. The Bank, together with other international agencies, particularly the UN and the IMF, has the ability to construct comparable data across countries. This is an important part of its role as a clearing house of development knowledge.

2. The Bank's policy work itself generates opportunities of learning and creating knowledge. When new programs are introduced with funding from the Bank, there is the opportunity to design an appropriate evaluation whose results become part of the knowledge base. While these evaluations could be done by outside agencies or researchers—and perhaps sometimes should be so done—the evaluation needs to be set up at an early stage as a routine part of Bank operations, which is something that could not easily be done by outsiders in a timely and efficient way. Outsiders are also unlikely to have the long run relationships with the operational staff and member governments that can be built up by Bank researchers.

3. There is a great deal of useful research that the academic community does not supply, largely because doing useful research (as against high visibility research) is not necessarily rewarded in academia. The World Bank, by virtue of being the largest publicly funded producer of development research and one that, for obvious reasons, has a strong stake in useful research, is the natural candidate for taking the lead in providing these intellectual public goods. The failure of academics in this context covers both whole areas of work, as well as types of work within areas. On the former, academic research is often fickle, with a flare of new work in a field, followed by years of neglect. Yet these can be areas that are of vital, everyday importance to the Bank and its members. Currently, there is very little frontline academic work being done by economists in such important areas as urban economics, transportation, climate change, and infrastructure. A prime example of the second kind of failure, research that is unlikely to be done by top academics is replication and testing. The fact that some new

idea worked in one location gives us hope, especially if there are good theoretical reasons to suppose that it will generalize, but there is no guarantee that it will work elsewhere. Trying it out in multiple locations is the only way to check. Yet within academia replicating what someone else has already done, although widely practiced, is not done systematically, and it is perceived as derivative and unoriginal, and not highly valued.

4. Bank researchers are also likely to be those in the best position to apply an existing body of theory, or of prior experience, to a specific practical problem. We now know a lot, to take an example, about how to analyze different types of auctions: what does that accumulated body of knowledge tell us about how to auction the airwaves in country X? Similarly, countries facing a particular problem, for example the privatization of an airline, the reform of a pension system, or the construction of infrastructure, are usually keenly interested in the experience of other countries that have already faced similar problems. Bank researchers are uniquely well-placed to synthesize such information from both theory and practice, and to present balanced and accessible accounts to member countries. There are no incentives to undertake such work in academia.

5. A related role for Bank research is in measurement, and more generally, in “descriptive” research, which is research that primarily answers questions of the form “how are things going?” and “what happened?” Measuring poverty or changes in poverty, or calculating purchasing power parity exchange rates are an example of the first kind of work, while a detailed history of a particular project, or regular country monitoring, are examples of the latter. At its best, this kind of work is quite analytical,

because it involves finding the right descriptive tools to bring the most out of the data and organizing the data in a useful way relative to existing theories, but it is unlikely to be seen as frontier work by academics. Methods of measurement need constant updating as the world changes, and the evolution of the conceptual basis of such updating is often difficult and intellectually challenging. In our view, the recent relative academic neglect of measurement is very costly, since it is measurement that generates the facts that animate all other research, that dictate priorities, and that provoke policy thinking. The Bank, unlike academic departments, who need the approval of the rest of academia when deciding who to promote, can set its own standards for research, though even Bank research needs at least some certification through the journal publication process in order to be credible.

6. World Bank researchers should be the best critics of the institution and its policies. Radical heterodoxy has a habit of becoming the orthodoxy of tomorrow, and one of the most important roles of Bank research is the analysis and criticism of current Bank policies. The Bank gets a great deal of criticism from the outside, at least some of which is political motivated, ill-informed about actual Bank operations, and impervious to empirical evidence. As insiders, Bank researchers can do much better, and the future health of the organization depends on them being able to do so.

7. Finally, the World Bank produces a set of research products that only a policy-making organization would want to produce. These are pieces of synthetic research, which aim to draw out the key lessons from the current body of original research in that area. The

World Development Reports, and more generally what the Bank calls its flagship research documents, fall in to this category. This is a kind of research that only someone who is extremely familiar with cutting edge research on the subject could attempt to do, but it is not something that is well rewarded in academia.

Problems of doing research in the World Bank

According to its own documents, Bank research has four objectives: (1) To generate knowledge to guide the Bank's corporate strategies, policy advice, lending operations, and technical assistance; (2) To respond to the specific needs of Bank operations, including assessment of development progress in member countries; (3) To generate knowledge that is primarily a global public good serving the development community; and (4) To assist in developing indigenous research capacity in member countries. These objectives are strongly endorsed by the panel.

In principle, fulfilling these objectives should be an attractive program for economists who are interested in changing the world. Bank researchers work full time on humanity's most pressing problems. They have great opportunities to observe policy, to develop new ideas, to see them put into practice, and they often have direct access to high-level policymakers, whose decisions affect the lives of millions of people. While they may not have the latitude of academics to follow ideas purely for their intellectual appeal they, unlike academics, are guaranteed a constant supply of problems whose solution is of real practical importance. Certainly, these advantages have attracted some of the best economists of each generation to spend time at the Bank as Chief Economist, and there

are many other Bank researchers who revel in the opportunities to combine practical policy work with new academic research.

Yet there are also many practical difficulties in fulfilling the four objectives at once, and it is important to keep these clearly in mind when evaluating the strengths and weaknesses of the Bank's research record. In academia, researchers' prime responsibility is to goal (3), the generation of new knowledge, and there exists a system of peer-review journals and personnel reviews to evaluate their contributions. While they are also expected to develop new capacity, through teaching, their students are carefully selected by ability and prior training. Bank researchers are expected to do all of this, together with a good deal of operational work, and with counterparts who, in some countries, have very little training. Some operational staff see little value in any kind of research, and think most of what they see is irrelevant. In other cases, researchers offer relevant advice that is unwelcome, because it challenges or undermines current modalities. Meeting academic standards is almost certainly necessary to certify the quality of Bank work, on which its standing in the development community depends, but academic journals and their referees have very different objectives from the Bank. At times, it appears that there is little work that is both relevant to operational staff and attractive to the academic journals. So that when Bank staff manage, on a regular basis, to fulfill both sets of criteria, as they regularly do, it is important to recognize the magnitude of their accomplishment.

Writing in 1997, before he became the Vice President for Research, Nicholas Stern (and Francisco Ferreira, who is now in the research department in the Bank), were quite

blunt about the difficulties of doing research in this passage from their article, “The World Bank as “Intellectual Actor”:

“Researchers are not free to follow intellectual inspiration. They are under constraints of designated priorities and of an apparent need to be immediately useful to operations. Further there is the strong hierarchy and an atmosphere much more deferential than would be found in universities. Among researchers there is considerable concern with what superiors will think of conclusions reached, to the occasional detriment of whether an analysis is sound”.

To this we would add that the superiors themselves are sometimes under pressure from the Bank Presidency and elsewhere not to say things that go directly against the broad policy line that the Bank is espousing. Stern and Ferreira (1997) give the example of Bank research on debt rescheduling and other ways of relieving pressure on developing countries during the debt crisis years of the 1980s. They suggest that the reason why Bank research lagged behind research outside which had already turned against forcing developing countries to pay everything they owed, is that the Bank’s major shareholders were worried about hurting the banks in their own countries.

The fact that the Bank values its connections with country governments (often for the very important reason that this is what allows it to have policy influence) also makes it hard for Bank researchers to publish and publicize results from research that the government does not like. Moreover who gets involved in the research gets decided by country teams, and country teams presumably base their selection in part based on what they are looking for (which may be a particular answer or a particular researcher who they know and like working with, or

perhaps someone known for not rocking the boat). The person who does the research is therefore not necessarily the one best suited to the job.

Finally, the Bank is an enormous organization and there is a wide range of opinions on any one issue within the senior leadership of the Bank. Moreover, different people value different things from research: some value more the policy message, while others might care more about the contribution to knowledge. Anyone doing research inside the Bank is therefore playing to many audiences, not the least because under the current matrix structure, he or she is answerable to at least two people. One might imagine that this could be a major distraction for any researcher.

The problem of multiple audiences is even more serious in the case of the flagships, where the message has a clear policy slant and the world is listening. It would seem naïve to believe that the writing of these can be fully insulated from thoughts about how the Bank's major shareholders would react, or how it would be received in the media, among academics or in the wider development community.

We must also recognize that, for a combination of reasons, working in the World Bank has not been seen as the most desirable first job for newly-minted PhD's. Some might attribute this to a too-narrow theoretical focus in academic programs, as well as the socialization of students to follow their teachers into academia where they have the chance to be intellectual leaders. At the same time, salaries for economists in the World Bank have not risen as rapidly as have those of academic economists while, on the supply side, the Bank has been authorized to hire very few young economists in recent years.

Some of the most distinguished researchers in the Bank started their careers as professors, but were attracted to the Bank by the opportunities outlined above. Yet it seems that this once fertile source of talent has become much less important in recent years.

Implications for the evaluation

In looking at Bank research since 1998, we followed two main routes of enquiry. One was to talk to a wide range of knowledgeable people, inside and outside of the Bank. The other was to ask a group of co-evaluators to read and comment on nearly 200 research projects. In the main chapters below, we shall report on our findings. But it needs to be remembered that each commentator provides, at best, a partial view, often confined to one or other of the four objectives of Bank research. Taken individually, these views can be misleading, because they do not appreciate the breadth of tasks that are expected of Bank research.

Our evaluation found a great deal of truly excellent work. There are Bank researchers who are repeatedly sought out by their operational counterparts, whose long-term relationships with country partners are highly valued on both sides, and whose work has shaped international thinking on the most important issues in development. Yet we will also report many detailed critical comments, and we conclude with a number of recommendations that we believe would make Bank research more useful on all heads. But none of this should detract from our overall assessment that the Bank's economists have done a creditable job of delivering on the many, potentially inconsistent, demands made of them. We believe that research is of vital importance to the World Bank, and that Bank researchers have delivered value to the institution that is much larger than could be

reasonably expected from the very small share of the budget that they command. The “Knowledge Bank,” to be worthy of its name, must surely devote more than a fortieth of its budget to the creation and communication of knowledge.

Chapter 2. How and where research is done in the World Bank

The World Bank produces a large number of “analytical” pieces both for individual country clients and cross-country studies. What is commonly known as analytic and advisory work includes (1) economic and sector work (i.e., reports on individual countries), (2) technical assistance and (3) research. The World Bank distinguishes research from other analytical work in that research is designed to produce results with wide applicability across countries or sectors¹, while economic and sector work take the product of research and apply it to particular project or country settings. For FY2005, research² was 11 percent of the budget spent on analytic and advisory work, which is consistent with the historical experience.

This report focuses primarily on research according to this definition. Examples of the forms in which this research is published are listed in Table 1 below.

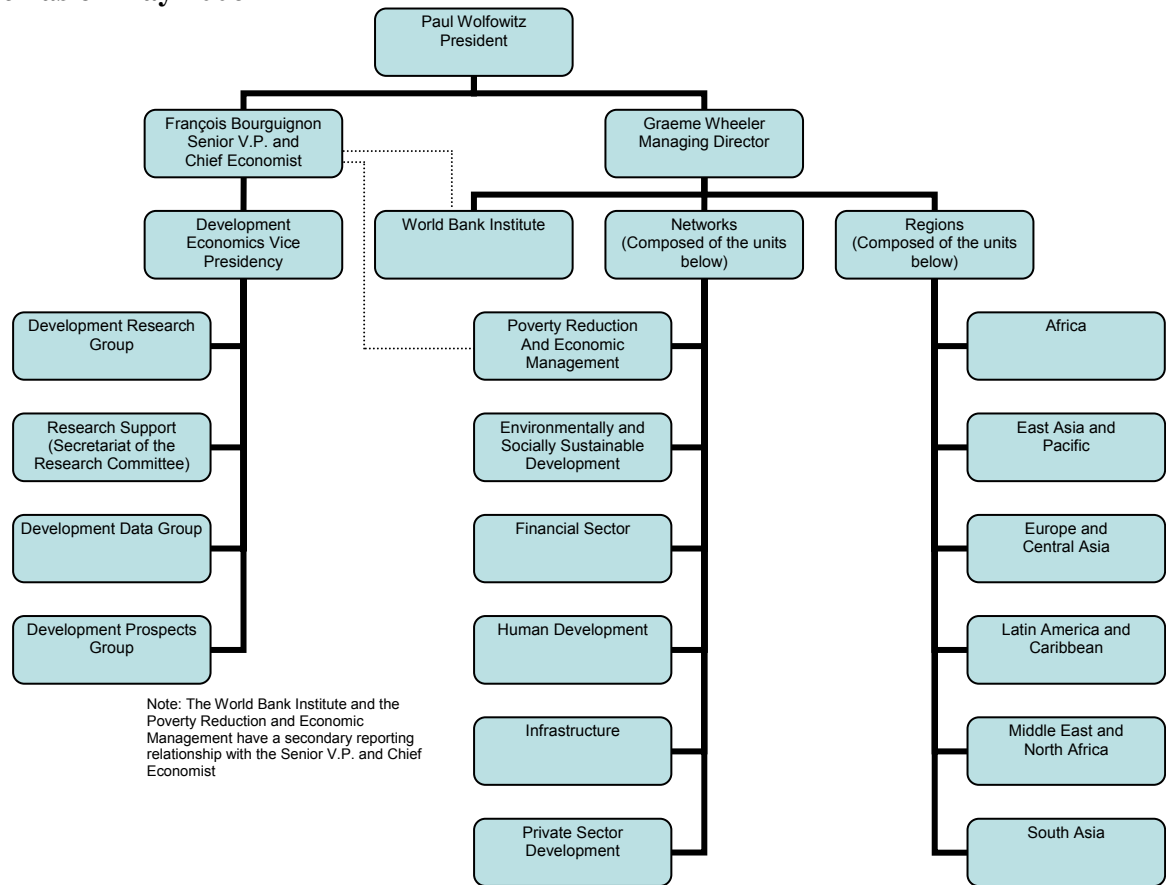
How Research is Managed and Organized

While most research takes place within the Development Economics Vice Presidency (DEC), research is produced throughout the World Bank, particularly in the regions, the networks, and in the World Bank Institute. The regions are responsible for the primary lending operations of the World Bank and the policy dialogue with governments, while the networks cover certain areas (e.g. infrastructure) and cut across all of the regions. Chart 1 details the organization of the units in the World Bank that produce research. The chart does not capture the “matrix” nature of the organization, according to which the

¹ Although this is admittedly a narrow definition of research

² This includes only projects in the World Bank which were classified as research in the accounting system by Managers. See Note 3 for limitations of this classification.

Chart 1: Organization Chart of Groups within the World Bank which produce research as of May 2006



economists, other than those in DEC, belong to *both* a network and a region.

Research is carried out under the broad direction of the Bank's Chief Economist. It is useful to think of this as being done according to three different modes: through the Research committee, within DEC, and within the regions. First, the Research Committee manages the allocation of the Research Support Budget. The Research Committee, chaired by the Chief Economist, is comprised by 19 managers from throughout the Bank, and distributes funds to research projects *throughout the Bank* in a competitive proposal process. In FY2005, \$6.5 million (approximately 26 percent of the total research

budget³) was distributed through the Research Committee. Second, the Development Economics Vice Presidency (DEC), particularly with help from the Development Research Group, helps determine what areas or issues DEC research staff will spend their time on. DEC reports directly to the Bank's Chief Economist. Third, the operational vice presidents of the regions, networks and the World Bank Institute also direct funds (at their discretion) towards research to address regional priorities or gaps in knowledge not covered by DEC or the Research Committee.

While the Bank's Chief Economist is Chairman of the Research Committee and as such has a great deal of influence over the Research Committee and DEC, he or she does not have direct control over funds used for research in the other vice presidencies. However, the Chief Economist does have a strong influence over what research is done outside of DEC, and regularly reviews many research outputs produced by the regions and networks. The regional chief economists⁴ also have a secondary reporting role⁵ to the Bank's Chief Economist. The Chief Economist also regularly consults with the operational vice presidents and regional economists when setting the research agenda.

In addition to own Bank budgetary resources, a significant portion of research done at the Bank is funded by trust funds (typically from the aid agencies of foreign governments, who are interested in supporting specific research areas) which Bank

³ The total research budget of \$25.3 in FY2005 includes all projects in the World Bank which were classified as research in the accounting system by Managers. However, due to the limitations of the accounting system, it may not include all of the costs associated with research outputs at the World Bank. It does not contain a significant share of the following costs: some network and regional flagships, staff and other costs of the Development Data Group and Development Prospects Group, the World Development Report, and cross-support among others. It does include costs associated with research dissemination.

⁴ The Human Development Network also has a chief economist.

⁵ The secondary reporting relationship gives the Chief Economist of the Bank limited influence over the regional and network chief economists through annual performance reviews and regular meetings.

researchers apply or bid for. In FY2005, trust funds provided approximately \$6.9 million (approximately 27 per cent of the total research budget⁶) in funding for Bank research.

Development Economics Vice Presidency (DEC)

The Development Research Group (DECRG) (the largest group within DEC) is the main research unit in the Bank, and currently has approximately 80 full-time staff researchers and 15 other long-term researchers, as well as around 30 support staff. The presence of full-time staff devoted to research makes the Development Research Group unique within the World Bank. The Research Group accounts for approximately half of the Bank's research outputs and collaborates with other researchers inside and outside of the Bank. In order to encourage researchers to maintain close ties with the operational side of the Bank, researchers have a target of 30 percent of their time that they must spend in cross-support to a region or network, typically providing analytical and advisory work to country clients.

Other groups within DEC include the Data Group (DECDG), the Prospects Group, (DECPG), and the Research Support Group, (DECRS), in addition to the communications staff in the front office. Research Support manages the Research Support Budget on behalf of the Research Committee, and has been the support group for this evaluation. The Data Group is the focal point for the Bank's work in data collection, global statistical work and monitoring, and assists other Bank researchers in data collection through tools such as the Development Data Platform. The Data Group also manages trust funds of approximately \$20 million for statistical capacity building. The

⁶ See footnote 3

Development Prospects Group produces projections of the global economy, and publishes them in flagship publications (described below).

Regions

The regions of the Bank divide the world up into Africa, East Asia and Pacific, Europe and Central Asia, Latin America and Caribbean, Middle East and North Africa, and South Asia. These regions sometimes conduct research in response to immediate problems arising in operational work, and occasionally undertake some long-term research. The regional chief economists meet regularly with the Bank's Chief Economist with whom they have a secondary reporting relationship. The frequency of research in the regions depends upon the management of the regions, and varies among the regions. However, the regions do not typically have staff devoted to research full-time, as is the case within DEC. The Latin American and Caribbean region is the most active region involved in research, with an annual budget of several million dollars. In contrast, the Europe and Central Asia region spends very little on research, apart from preparing some flagships. Recent flagship publications by the regions include *Inequality in Latin America*, *South Asia Pension Systems*, and *East Asia Decentralizes*.

Networks

The six networks were created in 1998. They are the Poverty Reduction and Economic Management Network, the Environmentally and Socially Sustainable Development Network, the Financial Sector Network, the Human Development Network, the Infrastructure Network, and the Private Sector Development. The role of the networks is to collaborate with and provide services to each of the regions in their area of expertise.

For example, the Human Development Network works with each of the regions to provide support for clients to improve human development. One manager explained to the panel that it is the role of DEC to create knowledge and the role of the networks to communicate it to the regions. In turn, the networks are supposed to help the regions find appropriate research resources in DEC. The organization of research within the networks varies from one to another, but is the responsibility of the Operational Vice President of each network. The networks typically do research in response to needs of the region.

The Poverty Reduction and Economic Management Network (PREM) is unique among the networks in that it has a secondary reporting relationship with the Senior Vice President and Chief Economist. PREM's objective is to integrate the Bank's poverty reduction efforts at the country level and provide policy advice to the Bank's clients during the formulation and implementation of policies and programs.

Publications of the networks have focused on issues such as systemic financial distress, pension support, or agricultural and the WTO.

World Bank Institute

The World Bank Institute is the Bank's capacity development arm, and provides learning programs and policy advice. The World Bank Institute also has an active publishing program, through which it maintains its standing in the academic world.

Types of Products

As shown in Table 1, the World Bank produces a number of research reports and publications aimed at a diverse clientele, ranging from academics, policymakers, and the general public. Although each research product may be read by any audience, the

primary audience of the research product is shown in Table 1. A brief discussion of each of these categories of research products follows below.

Table 1: The audience for World Bank research ranges from academics to the general public	
World Bank Research Product	Primary Audience
Journal Articles and Books	Academic
Policy Research Working Papers	Academic
Analytical Tools	Academic/Policymakers
Policy Research Reports	Academic/Policymakers
Data Products	Academic/Policymakers
Annual Bank Conference on Development Economics	Academic/Policymakers
Special "Flagship" Reports	Policymakers/General Public
World Development Report	Policymakers/General Public

Journal Articles

While researchers within DEC are expected to publish two articles per year in a peer-reviewed journal, Bank staff from regions and networks also publish in academic journals. Between FY 1998-2005, Bank staff produced more than 2,000 articles in peer-reviewed journals. In addition, the World Bank publishes two peer-reviewed research journals, the *World Bank Research Observer*, and the *World Bank Economic Review*.

The *World Bank Research Observer* is intended for people with a professional interest in development, while the *World Bank Economic Review* specializes in empirical analysis of development policy. These journals primarily publish papers written by World Bank staff or as a result of World Bank supported research, but they also accept submissions by outside researchers on a limited basis. The *World Bank Economic Review* has had an open submissions policy since 2004.

Policy Research Working Papers

The Policy Research Working Paper Series was established in 1988 to encourage the exchange of ideas on development issues and to disseminate the findings of work in

progress. Approximately 330 working papers are published each year, with nearly 4,100 published since inception. According to the Social Science Research Network, these working papers are among the most widely used economics papers downloaded from the internet, and may be the most widely used research output from the World Bank used by academics. Many are also targeted at policymakers.

Analytical Tools

In addition to publications, Bank researchers also devise new methodologies, approaches and tools to analyze development and policy problems which are used by researchers and policymakers. These include measures of poverty scenarios, poverty counts; economic models (e.g. computable general equilibrium), poverty maps, and surveys such as the public expenditure tracking surveys or the investment climate assessment.

Policy Research Reports

The Policy Research Reports are intended for a broad audience, and summarize research on development policy issues carried out by World Bank staff and other researchers on a particular topic. Approximately one per year is published. They are intended to provoke debate in both the academic and development communities on appropriate public policy objectives and instruments for developing economies. These reports are produced by the Development Research Group. Examples of recently published Policy Research Reports include *Reforming Infrastructure: Privatization, Regulation, and Competition*, and *Land Policies for Growth and Poverty Reduction*.

Data Products

The Development Data Group and Development Research Group collect and compile datasets which are used by policy analysts and researchers. The most prominent dataset published is the *World Development Indicators*, which aggregates economic, environmental, and social data on over 150 countries. Other data products the World Bank distributes cover the areas of poverty, governance, finance, environment and health, international comparisons, and investment climate. Some databases are offered free of charge and some on an annual subscription basis.

Annual Bank Conference on Development Economics

Established by the Research Committee for the presentation and discussion of new knowledge about development, this is the Bank's best known conference series. It is held twice a year, with recent conferences held in St. Petersburg and Tokyo.

Special Flagship Reports

On an as-needed basis, the World Bank publishes special flagship reports which cover a particular region or a topic in-depth. Many are published by the regions and networks, often in collaboration with researchers from DEC. Recent reports include *Lessons from NAFTA*, *Unlocking the Employment Challenge in MENA*, and *Old Age Income Support in the 21st Century*. One annual report which has been a top seller is *Doing Business*, provides objective measures of business regulations and their enforcement over 155 countries. *Doing Business* is published by the Private Sector Development network. Other prominent annual reports aimed at policymakers include *Global Economic*

Prospects and the *Global Development Finance* produced in the Development Prospects Group, and the *Global Monitoring Report*, produced in DEC.

World Development Reports

The World Development Report, an annual report aimed at the general development community, provides in-depth analysis of a specific aspect of development. This report is written by a team of Bank staff and outside consultants, under the general supervision of the Chief Economist. The team also produces background papers on the issues discussed in the report. Recent reports include *Equity and Development*, and *A Better Investment Climate for Everyone*. According to Development Research Support, the World Development Reports are the Bank's best-known contribution to knowledge about development.

The Bank's Research Portfolio

Table 2: World Bank Research covers a wide range of topics

Research Topic	Abstracts from FY2001- FY2004	
	Count	%
Poverty and Social Development	73	16%
International Economics	60	13%
Environment	47	11%
Infrastructure and Urban Development	47	11%
Governance and Public Sector Management	41	9%
Agriculture and Rural Development	40	9%
Education, Labor and Employment	40	9%
Domestic Finance	33	7%
Health and Population	28	6%
Industry, Investment Climate and Private Sector Development	19	4%
Macroeconomics and Growth	17	4%
Total	445	100%

Source: *Abstracts of Current Studies 2001, 2002-03, 2004*

Table 2 details the number of research projects initiated between FY 2001-FY 2004, and illustrates the breadth of the World Bank research portfolio. Approximately half of the projects focus on four areas: Poverty and Social Development, International Economics, Environment, and Infrastructure and Urban Development.

The Bank's current research agenda priorities fall into three groups: growth and equity, global issues and evaluation. Growth and equity include issues such as investment climate, public services and public goods, and progress towards the Millennium Development Goals. Global issues include international trade, migration and security and development. Evaluation involves an effort to systematically compare the effectiveness of specific interventions in different settings and follow alternative designs.

The Development Research Group's current work is grouped into seven teams: Finance, Growth and Investment, Human Development and Public Services, Infrastructure and Environment, Poverty, Rural Development, and Trade and International Integration. In addition, the Director's Office undertakes specialized research programs, such as one entitled "East Asia's Future Prospects". The work of the seven teams in the Development Research Group is described briefly below:

Finance

This group focuses on understanding how financial systems contribute to economic development and poverty reduction. Their research also attempts to identify policies that improve the effectiveness, stability and reach of financial systems. Examples of current research include work on bank privatizations, small and medium enterprises and access to finance.

Growth and Investment

The research in this group aims to identify and improve policies and reform strategies which are conducive to sustained growth. In addition, the research aims to understand the factors behind the diversity in aggregate economic performance across countries, as well as their varying responses to policy and institutional changes. Examples of current research projects include “Inequality and Discrimination”, and “Regulation, Institutions and Growth”.

Human Development and Public Services

The research in this team aims for a deeper understanding of the factors affecting human development in developing countries, improve the analysis of service delivery and examine the effectiveness of aid. Examples of current projects include “Teacher and Health Worker Absenteeism”, “Quality of Health Care”, and “Child Growth, Income Shocks, and Government Programs”.

Infrastructure and Environment

The team’s research looks at policies and institutions that are intended to reduce environmental damage, improve regional strategies for sustainable development and improve the contribution of urban development to reducing poverty. Recent research projects include “Indoor Air Pollution” and “Traffic Fatalities and Economic Growth”.

Poverty

This poverty team's research has two major objectives. The first is to improve current data and methods of poverty and inequality analysis, which includes producing new household-level data and developing "poverty maps". The second is to improve data to better understand the economic and social processes determining the extent of poverty and inequality and to assess the effectiveness of specific policies in reducing poverty. Recent research papers include "Lessons from China's Progress against Poverty" and "Impact Evaluation of Antipoverty Programs".

Rural Development

This research program focuses on understanding the factors and institutions that generate and perpetuate rural poverty, and an assessment of policies and interventions designed to support poor people in rural areas in improving their lives. Recent research highlights include work on water resources and land markets.

Trade and International Integration

This research team attempts to better understand the role of goods, services, and factors of production in economic development, and also to assess and create policies to enhance the gains from integration. Recent research projects include "Bolstering the Case for Agricultural Trade Reforms" and "Migration: Brain Drains and Brain Gains".

Chapter 3. Assessing the Quality of World Bank Research

The panel and a distinguished group of academic co-evaluators read a sample of all research from 1998 to 2005, including research from throughout the Bank. Much of what we read was of very high quality, was directed toward issues that are of great importance to the Bank, and was executed to the highest standards of the profession. There are things that the Bank does well such as creating and organizing knowledge based on operational experience and on its unequalled access to countries and country evidence. Its research has had a major effect on the way that development issues are discussed by practitioners, policymakers, and academics. It generates a steady flow of papers, a few of which are published in the most prestigious journals in economics⁷ and other areas. Bank research has established the Bank as a major intellectual force in the development community.

Bank researchers are also crucial creators and disseminators of data that are widely used by the countries as well as by the development and academic communities. They generate research that is tailored to operational needs, and that evaluates and sometimes challenges operational practices and outcomes. They have managed to maintain a strong research presence in some important operational areas where there is little or no outside academic research to support them, and they have taken the lead in working on topics, such as civil wars, aid effectiveness, doctor and teacher absenteeism, or pollution in developing countries that other researchers have unduly neglected.

⁷ There are 3,798 papers and books by Bank staff and consultants in the (extensive but almost certainly incomplete) Bibliography in Annex A. Of those, 35 appeared in the three top general interest journals in economics, the *American Economic Review* (21, about half of which are short papers, comments, etc.), the *Journal of Political Economy* (5), and the *Quarterly Journal of Economics* (10), as well as two papers in *Econometrica* and seven in the *Journal of Finance*.

At the same time, we found a number of deficiencies. Alongside the excellent work, there is a great deal of research that is undistinguished and not well-directed either to academic or policy concerns. We emphasize that this judgment is not based on citation counts, or on the number of papers published in leading journals. Bank work is often well-cited, and there are many important and innovative Bank papers in good journals. The judgment comes from reading papers that seemed to contribute little that would be useful either to policymakers or to academics. We think there is more such work than is to be expected from an admittedly high risk activity such as research and even recognizing the unique constraints on Bank researchers. The quality of execution does not always match the importance and the relevance of the topic, and is often unacceptably far behind best-practice methods.

A small fraction of prominent Bank research is technically flawed and in some cases strong policy positions have been supported by such (non) evidence. The panel fully appreciates the need to take positions before all of the evidence is in, and recognizes that the Bank must often aggressively defend its own policies. But putting too much weight on preliminary or flawed work could expose the Bank to charges that its research is tailored or selected to support its predetermined positions, and the panel believes that, in some cases, the Bank proselytized selected new work in major policy speeches and publications, without appropriate caveats on its reliability. We believe that this happened with some of the Bank's work on aid effectiveness, which we discuss at length below.

New research methods have sometimes found their way into country assistance and country policy without adequate evaluation. One example that we discuss at some length is the innovative and potentially important work by Bank researchers on the estimation of

poverty for small areas. This is an important case in its own right, but it also illustrates broader issues of statistical and econometric practice in the Bank's research work.

The panel found that non-Bank advice and consultation is done on a largely haphazard basis and largely left to individual researchers, so that some of the very best work and some of the very worst work that we reviewed were written jointly with outside consultants. At the same time, there is remarkably little work co-authored by non-Bank researchers from developing countries. Such lack of involvement may reduce the Bank's contribution to the development of research capacity in borrowing countries and inhibit the influence of Bank research where it is most needed. Although there are many other avenues of dissemination, our impression is that the Bank has not always been as successful as influencing the debate as other, smaller but nimbler research groups with much better and more user-friendly websites. Some of the "flagship" publications, both from the research group, and from other parts of the Bank have been notably successful at summarizing and conveying the products of Bank and outside research. Even so, the Bank produces too many long (and sometimes unreadable) book-length reports that are ostensibly directed at policymakers, but seem very unlikely to be read by them. The most important of the Bank's flagships, the series of *World Development Reports*, are widely read, and sometimes affect the international debate, as well as the debate within the institution. At the same time, they absorb a large fraction, around ten per cent, of resources within the Research Group.

We found differences in the quality of research depending on whether or not it was carried out by researchers in the Development Economics Research group (DECRG), or

elsewhere in the Bank. Our evaluators generally rated research from DECRG ahead of non-DEC research, whether from the regions or the networks. An exception was the “flagship” reports from outside of DEC, many of which were thought to be of high quality. DEC research tended to be methodologically stronger than non-DEC research. More generally, and over all of the research that was reviewed, we found that Bank work scored highly on the importance of the topic, but less highly on execution, particularly on the appropriate use of methodology. One criticism that was made repeatedly is that research tended to jump to policy conclusions that were not well-supported by the evidence. An appendix to this chapter presents the quantitative analysis on which these conclusions are based; the qualitative evidence is reported in the main text below.

How the evaluation was organized

In conjunction with the research management of the Bank, the panel approached a group of twenty-six distinguished researchers, mostly academics, who were asked to read a selection of Bank research within their fields of expertise. All of these scholars are leaders in their field, and we were fortunate enough to attract some of the superstars of their cohort. As was the case with the panel, this group covered a range of nationalities, and included researchers who work (or have worked) in South America, Asia, and Africa as well as in the United States and Europe. The majority of the co-evaluators had worked as consultants for the Bank at some time in their careers, one or two for extended periods in operations as well as in research. Several had never worked for the Bank in any capacity. Given the Bank’s importance in development research, and the ubiquity of at least occasional Bank consultation among academics, it was felt that any loss of

objectivity was more than compensated by the competence and knowledge that came with previous contact with the Bank. It would also have been close to impossible to find a group of evaluators who were professionally competent but who had never worked for the Bank. The names, nationalities, and affiliations of the evaluators are given in an appendix to the report. We were extremely fortunate in the people who agreed to work with us as co-evaluators; we find it hard to imagine a group of evaluators who would be more distinguished or more qualified to evaluate the quality of development research.

We were charged with evaluating all Bank research over the period from 1998 to 2005. A complete list of research *projects* over this period was prepared by the Bank's research support administration, and using that list as a frame, we drew a stratified random sample of 180 projects for evaluation. Projects were stratified by size, by DEC and non-DEC, by size, and by date of starting. To the random sample was added the complete list of World Development Reports over this period, which were read by the panel members themselves, as well as (a majority of) the regional and network flagship publications, most of which were assigned to existing evaluators, although in three cases, we enrolled new evaluators with the relevant expertise. We also asked the Bank's research director to nominate a group of "must read" outstanding papers or books from DEC, some of which were not in the sample, and these were also assigned to evaluators. The panel felt that it was more important not to miss the best work than to maintain the purity of the random sample, even at some risk of some bias in favor of DEC research. Even so, our own and the evaluators' reading did not always rate these papers highly.

The sampling was necessitated by the sheer volume of the work; the research universe comprised projects which had generated a total of nearly 4,000 databases,

papers, and books or book length manuscripts. Inevitably, there was some arbitrariness in the sampling design. Particular items of research are not always readily assignable to particular projects, and research funds, in the Bank as in academia, are to some extent fungible between projects, so that it proved impossible to draw up a complete correspondence between nominal funding sources and outputs. This meant that the panel was unable to make any assessment of value for money for any particular research project, although that was part of our original remit. We believe that such an assessment is extremely difficult under any circumstances, but it is certainly cannot be adequately done given the current Bank accounting and monitoring procedures. Even so, there were some specific projects where it was possible to provide some comments on value for money, and evaluators occasionally did so.

Evaluators were assigned a varying number of projects, depending on the amount of research in their field, but all did a very large amount of reading. A typical assignment covered ten projects, or thirty or so papers and books. Each reader was asked to provide detailed assessments and scores on various criteria for each project or stand-alone piece of research. They were asked to assess the extent to which the research they read contributed to achieving the two main objectives of World Bank research, the generation of new knowledge on development as well as contributing to broadening the understanding of development policy. In addition to their detailed project reports, evaluators were asked to provide an overall assessment of Bank research in their area of expertise, and in this capacity they were encouraged to use their knowledge beyond the work that they had been assigned. It is important to note that, while they were expected to report on the technical and academic value of what they read, they were not asked to

evaluate the research on purely academic grounds, although there was certainly some heterogeneity of approach. But as is clear from the majority of the reports, evaluators were much concerned with policy relevance, and recognized that Bank research is conducted with different objectives and under different constraints than is academic research. Indeed some commentators explicitly discussed what might reasonably be expected of Bank researchers who are simultaneously expected to meet standards of both academic and policy relevance, standards that are sometimes in conflict.

The panel is extremely grateful to the evaluators for the extraordinary volume of work that they did, and for the thoroughness and thoughtfulness with which they did it.

The discussion that follows is based on the evaluators' comments, which are presented in a separate volume that is a supplement to this report. The panel also undertook a great deal of reading of its own, and also held a one-day meeting with the all but five of the evaluators, during which there was a wide-ranging exchange of views. The comments that follow are based on all of this evidence, and will occasionally differ in detail or in emphasis from the evaluators' views, with which we occasionally disagree. In this chapter, our main emphasis is in reporting and summarizing what we learned about the characteristics of Bank research. Why it is what it is, and recommendations for the change, are taken up in Chapter 6.

We emphasize again that our evaluation was did not involve a complete reading of Bank books, papers, and reports. There was simply too much output (nearly 3,800 pieces are listed in the Bibliography) to make a complete evaluation feasible. In consequence, there are undoubtedly important pieces of research that we missed, while in other cases, our results might have been affected by the idiosyncratic views of specific evaluators, or

of the allocation of projects to readers. These caveats should be born in mind, particularly in reading the discussion of particular projects, though we do not believe that they had an important effect on our overall conclusions.

We should also note the special position of two groups within DEC, the DEC data group, DECDG, and the DEC prospects group, DECPG. Both of these groups were subject to separate evaluations that ran concurrently with our own. In consequence, the panel did not focus *directly* on either group. In the case of DECPG, this meant that their activities are largely absent from this report. The DEC data group, by contrast, will appear frequently in our discussion, if only because it plays such an important part in the research activities of the Bank, which are our primary focus, and because its work and relations with researchers were repeatedly raised in the evaluations. Even so, there are important activities of DECDG that are not discussed here; for example, the collection and dissemination of important information on debt and on national accounts, and the support for monitoring the Millennium Development Goals, as well as international statistical cooperation and capacity building. These topics were left to the separate evaluation, so that a full appreciation of DECDG requires examination of both reports. In what follows, DECDG will be referred to on many separate occasions, and in different contexts, though we try to bring together the various recommendations in the final chapter.

A note on citation analyses

Citation analysis has become a ubiquitous tool of quality evaluation, and we will occasionally make reference to citation numbers in the rest of this report. However, the

panel believes that it would be a mistake to put much weight on citations and that they are of very limited value for this kind of evaluation.

Citation counts are invaluable for assessing the long-term impact on subsequent work of individual scholars, or collections of individuals, as in academic departments. They are much less useful in the short run. The period we are assessing here is 1998 to 2005, and there simply has not been time for the citation record to accumulate. The most reliable citation databases, such as Thomson's ISI web of knowledge, allow careful control of the search process, but exclude citations to working papers. Given that top journals in economics take three years or more to publish even the best papers, this exclusion severely limits the usefulness of these data for our purposes. Google Scholar does reference working papers, and is in many ways much more useful, but it is much harder to control, for example by limiting the search in various ways, and its search universe is often unclear.

But we have a much more positive reason for not using citations. Our evaluators represent the very best in contemporary research in economic development, and they did what the databases cannot do, which is to read the work. Their reviews, on which this report is based, are careful, detailed, and deeply informed. They both commend and criticize. These evaluations have little to do with citation counts, and indeed, in some of the cases that come in for the sharpest criticism, the work has been heavily cited. But this is exactly the point of an evaluation like the current one. One of our most important tasks is to draw attention to good, policy-relevant work that is not getting the notice or credit that it deserves. But even that is less important than identifying cases where the Bank, its

readers, and its clients are relying on Bank ideas or methods that are widely influential, heavily used, but deeply flawed.

Evaluation results: an overview

Highlights: Bank research at its best

The panel is enormously impressed by the best of the Bank's research. There is a great deal of work that meets the highest academic standards of originality and technique, and Bank work is frequently published in the top academic journals, such as the *American Economic Review* (21 papers over the review period, 10 of which are refereed research papers), the *Quarterly Journal of Economics* (9 papers), the *Journal of Political Economy* (5), and the *Journal of Finance* (10), as well as in relevant field journals, particularly the *Journal of Development Economics* (60), and, where relevant, in a number of prominent non-economics journals. At the same time, this work is concerned with topics that are vitally relevant to the Bank's mission, issues such as poverty measurement, evaluating the effects of Bank projects on people's wellbeing, whether aid works, on the extent to which growth alleviates poverty, on corruption, the environment, civil war, trade arrangements, decentralization, and much else.

The evaluators and the panel identified a number of specific projects which deserved particular praise. The panel was particularly impressed by the program on service delivery, and on the behavior of teachers and of doctors. The survey of teacher absenteeism that has just been published in the *Journal of Economic Perspectives* documents the extraordinary degree of absenteeism by teachers in countries around the developing world. Similar Bank work on doctors and health providers, who show

similarly high rates of absenteeism, has provided the documentation for a widespread problem that was not well understood, either by researchers, or by policymakers in the countries themselves. Documentation, by itself, is an important first step in addressing the issue.

On the behavior of doctors, there is also an impressive set of Indian studies by Jeffrey Hammer and Jishnu Das that are casting light on the competence of doctors, and on the different incentives and behaviors of private and public doctors in India, with neither group providing good service, although each fails in different ways and for different reasons. Most of death and disease in poor countries is neither attributable to the absence of appropriate medicines, nor to a lack of an appropriate method of treatment, but comes, among other things, from failures of health delivery, so that work like this, which adds to our understanding of the mechanisms of failure, is of great importance.

Another notable piece of Bank research is based on a paper by Deon Filmer and Lant Pritchett that was published in *Demography* in 2001. Filmer and Pritchett had the clever idea of using information on household ownership of durable goods to construct a rough and ready measure of household wealth. This is useful because many countries around the world have collected data using Demographic and Health Surveys (DHSS), a more or less standardized survey that is funded by USAID. These surveys do not collect any good information on economic status, but they do collect ownership of durable goods, so the Filmer and Pritchett index provides an admittedly crude but valuable indicator of economic status. In their original paper, they used this to show that children in families with lower wealth (by their measure) are less likely to go to school, but the most widespread application of their methods has been to health. Before they proposed their

index, there was essentially no way of investigating how mortality rates vary by socioeconomic status in poor countries. Since then, an enormous industry has sprung up documenting health inequalities, not only in the Bank, but in other agencies, such as the WHO, as well as in academic papers. Although the method is probably overused, and many of those who use it do not understand its limitations—for example that quintiles defined by the index are not the same thing as income quintiles—there is no challenging the enormous influence of the methodology. While the idea is straightforward, this is a good example of work that probably would not have been done outside the Bank, if only because academic standards undervalue the importance of description and documentation, particularly of standards of living among the poor. And indeed, our own evaluator dismisses this paper in his report⁸.

The panel and the evaluators were also impressed by the extent to which Bank researchers are seriously experimenting with project evaluation using randomized controlled trials. Not all of the work that we saw was successful, and perhaps the most cited and highest profile of the Bank funded studies is the pseudo-randomized study of helminthic infections in Kenya by two bank consultants, Michael Kremer and Edward Miguel, and published in *Econometrica*. We have no doubt that there are more opportunities for this kind of work in the lending portfolio, and it appears to have taken root at the Bank.

Bank researchers have also generated a vast amount of data, much of which represents innovative thinking and incorporates new design, including the long-standing and successful Living Standards Measurement Surveys (LSMS), as well as new and

⁸ Unlike the panel, and in disagreement with them, he also finds little value in the Das and Hammer work discussed in the previous paragraph.

widely-used databases, such as the Doing Business surveys, the Investment Climate surveys, and the Business Environment and Economic Performance surveys (BEEPS) in the transition countries. The impact of these new surveys is noted by our evaluators. Antoinette Schoar (MIT) writes that “Doing Business is one of the most influential research initiatives that the IFC and the World Bank have ever undertaken. It has put the focus on improving the efficiency of government policy and ignited a vigorous discussion in emerging markets.” She notes that entering the generic term “doing business” into Google leads to hits, the first three of which are to the Bank’s Doing Business website.

Francesco Caselli (London School of Economics) writes of the Investment Climate Surveys that “I have nothing but praise for this initiative.” He notes that it is entirely consistent with, and may have contributed to a new emphasis in macroeconomic research on firm dynamics; “the profession increasingly recognizes that it is difficult to think of aggregate investment, technological progress, employment, and growth without understanding the details of firm-level decision making, and particularly the constraints that firms face. This is an area where the Bank is not only abreast of current initiatives in macroeconomics and growth, but is providing leadership.” Both the investment climate surveys and the doing business surveys have excellent websites that allow wide public access to the information that facilitate the use of the knowledge that the Bank has created.

The Bank’s poverty research group originated the dollar-a-day poverty count around 1990, and halving the poverty count is now the first of the Millennium Development Goals. The methodology underlying these estimates, and the scorekeeping in the future, is carried out by the poverty group in the Bank. The results are published in the World

Development Indicators, and on a poverty monitoring website that also permits anyone in the world to inspect the information underlying the poverty counts, and indeed to provide alternative counts based on different assumptions. Dollar a day poverty counts can also be generated for a wide range of countries.

The Bank's data group, DECDG, which is lodged within the Development Economics Center, maintains and disseminates (at a charge) what is almost certainly the most heavily used database by development researchers and practitioners, the World Development Indicators. While much of this activity is data collation from other original sources, DECDG is increasingly moving into the business of data production. Over many years, it has produced important information on child and infant mortality that is to some extent independent of the data produced by the United Nations. Most importantly, the International Price Comparison Project has recently moved into the World Bank, and the DECDG is currently undertaking the enormous worldwide data collection effort that will produce the next set of purchasing power parity price indexes. This effort, which began in the University of Pennsylvania in the 1970s, and has been previously presented through a series of Penn World Tables, is an undertaking of the greatest importance for any and all attempts to measure economic growth, living standards, and poverty around the world. It is a global public good of the first magnitude, and central to the monitoring of poverty and development in the world.

Perhaps the most cited and influential of all of the Bank's research outputs is William Easterly's book, *The elusive quest for growth*, published in 2001, and which at the time of writing of this report, has accumulated 673 cites according to Google Scholar. This book, together with Sachs' book *An end to poverty*, and Easterly's new book, *The*

white man's burden, have dominated recent discussion on the effectiveness of foreign aid over the last five years, and have been hugely influential on development thinking everywhere. Easterly's (first) book clearly could not have been written without his long experience in the Bank, and even if it challenges much of the way that the Bank does its business, the panel does not believe that it is any the less important, or any less of a signal achievement of the Bank's research. Indeed, as we argued in Chapter 1, such fundamental challenges are one of the most important roles to be played by a research group in any such organization.

Our evaluator was also impressed by the Bank's project on the "Greening of Industry," which Geoffrey Heal described as "a truly first-rate project," which "shows that there are many mechanisms that work to control pollution even in the absence of formal pollution-control legislation, or in situations where such legislation exists but is not implemented." This program has been associated with a system of environmental rating measures that allow interpretation of the significance of emissions, as well as providing public disclosure of them; these measures have been implemented in a number of countries, including Indonesia, the Philippines, China, Vietnam, and India, where pilot programs have indicated significant reductions in pollution.

Important topics, but with serious shortcomings in execution and conclusions

While the panel and the evaluators found a great deal to praise, much of the research read by the evaluators was seen as undistinguished, and not well-addressed to any particular audience, either of academics or of policymakers. Some of this work contains flaws of one kind or another, some of which are equally problematic for academic and policy

work, and we will summarize some of the widespread concerns below. However, the evaluators and the panel were less concerned with flaws in the lower level and non-influential work than with flaws in some of the higher profile papers, including some of those that were praised on grounds of widespread influence and relevance to the Bank's mission. So we begin with these.

Globalization, aid, and poverty

The influence of Easterly's work has been equaled only by a series of papers and reports that use cross-country evidence to study how globalization affects poverty in countries with and without good policies. The paper by Craig Burnside and David Dollar published in the *American Economic Review* in 2000, "Aid, policies, and growth," has currently 743 cites according to Google Scholar. Contrary to Easterly's arguments, this paper, which argues that aid is effective in countries with good policies, has become the orthodoxy for those who are in favor of aid, and is cited in many prominent Bank documents. Dollar's widely cited (893 cites on GS) paper with Aart Kraay on "Growth is good for the poor," needs neither abstract nor summary. Another paper by Dollar and Kraay, in the *Economic Journal* in 2004, argues that countries that used large tariff cuts to open their trade to the beneficial effects of globalization have seen more poverty reduction than those that have not. Many of these arguments are brought together in a 2001 Policy Research Report on *Globalization, growth, and poverty* written by Dollar and Paul Collier. All of this work has had an enormous influence on the intellectual debates about globalization and poverty reduction and, to many around the world, it is

seen as defining the World Bank's position on these issues, as well as establishing the Bank's intellectual leadership in the globalization debate.

The panel agrees that this provocative research program has set out some stimulating research questions, and applauds the Bank's initial efforts. At the same time, however, we see a serious failure in the checks and balances within the system that has led to Bank to repeatedly trumpet these early empirical results without recognizing their fragile and tentative nature. As we shall argue, much of this line of research appears to have such deep flaws that, at present, the results cannot be regarded as remotely reliable, much as one might want to believe the results. There is a deeper problem here than simply a wrong assessment of provocative new research results. The problem is that in major Bank policy speeches and publications, it proselytized the new work without appropriate caveats on its reliability. Unfortunately, as one reads the research more carefully, and as new results come in, it is becoming clear that the Bank seriously over-reached in prematurely putting its globalization, aid and poverty publications on a pedestal. Nor has it corrected itself to this day. We wish to emphasize that we, too, believe that countries with good policies and institutions are far more likely to benefit from aid than, say, countries with deep corruption and poor governance where aid can delay reform rather than enhancing it. There is a strong theoretical presumption in favor of this commonsense dictum. However, it is very unclear empirically where the line can be drawn, or which policies matter, and in our view, the jury is very much still out on any quantitative assessment of the issue. Nor does the panel challenge the Bank's need to mount strong arguments in favor of its policies. Our problems are with the way that Bank research was used in the process, given the great credibility attached to what the Bank says.

Some of the Bank's research on globalization, aid, policy, and poverty, was read by Francesco Caselli (LSE). He reviewed the Bank's 1998 Policy Research Report *Assessing Aid*, written by Dollar and Lant Pritchett and which makes heavy use of the Burnside and Dollar argument that aid reduces poverty in countries with good policies. (The Burnside and Dollar results were then available as a 1997 Working Paper.) The argument is packaged for the broader policy community, and is used to argue for focusing aid on countries where there is both poverty and good policy. But as Caselli notes, subsequent studies, including a state-of-the-art study from the IMF that, although still within the tradition of cross-country regression work, is methodologically stronger than earlier work, have shown that the Burnside and Dollar results are not robust. It is possible to argue that the authors of *Assessing Aid* were simply unlucky, and that they could not know that the ground on which they chose to take their stand was so deeply undermined. But the panel takes a somewhat different view. There is at the very least a good argument that it should have been clear from the outset that the evidence could not bear the weight that was placed by it in the arguments about, and justification for, Bank policy.

In spite of having been published in the *American Economic Review*, the Burnside and Dollar paper is unconvincing. The analysis uses an index of policy that combines the government surplus, the inflation rate, and an openness measure, at least two of which are measures of *outcomes*, not of policies (as is indeed recognized in later work by Collier and Dollar). It is also clear from the way in which this index is constructed that the results are not robust; attempts to work with all three measures fail, as does a principal components index, and the final index is constructed using a regression of growth on

policy that is at best arbitrary, and at worst appears to be inconsistent with the main equation of interest. But this issue is dwarfed by the specter that haunts all of this literature, that external aid is not only a determinant of economic performance, but is determined by it. Burnside and Dollar conform to the previous (and subsequent) literature by using an instrumental variable technique, but this is a chicken-and-egg problem that is not readily resolved by mechanical means. In particular, it would require an unusually generous suspension of disbelief (even for cross-country regression analysis) to accept the identification assumption that the size of a country's population, the (one period lag) of the share of its imports that come in the form of arms, and whether or not it is Egypt have no effect whatever on their economic growth rates *except in so far as they affect its receipts of foreign aid*. Again, we are not arguing that the Burnside and Dollar paper is weaker than most of the literature on aid effectiveness, but we *are* arguing that its results provide only the weakest of evidence for their central contention, that aid is effective when policies are sound.

The Bank did not appear to recognize the weakness of this evidence. Not only did it form the basis for the PRR *Assessing Aid*, but its results were built upon in a series of papers by Collier and Dollar that were published between 2001 and 2004 in the *Economic Journal*, in the *European Economic Review*, and in *World Development*. These papers use an arguably improved indicator of the quality of economic policy (derived as an average of scores by Bank staff on a number of dimensions) but they make no attempt to deal with the chicken and egg problem, arguing that because Burnside and Dollar's results were very much the same whether or not they used instrumental variables, there is no need to worry. So their results, which go into the further Bank documents cited below,

are derived by ordinary least squares regression. It is not clear why such an argument would hold for a different data set and different variables but, in any case, it is founded on the scarcely credible assertion that the identification assumption in Burnside and Dollar are valid. The Bank then built on this second round of work, giving it extended prominence in a 2002 book *The Case for Aid*, which brings together a speech by President Wolfensohn, a paper summarizing the Monterrey consensus by Chief Economist Nicholas Stern, and a monograph-length substantial analysis “The role and effectiveness of development assistance,” by Ian Goldin, Halsey Rogers, and Stern, written for and presented at the Monterrey conference. The monograph brings together evidence from a number of sources, but it contains detailed calculations of the effectiveness of current Bank aid, now directed to poor countries with good policies, as opposed to aid that is not so directed, including some of the Bank’s own previous lending.

We think that the Bank was unwise to place so much weight on one paper whose evidence is so unconvincing. At the same time, there was other work being done by Bank researchers, particularly by Easterly, that did not find aid effective, even conditional on good policies. That work shares many of the problems that are inherent in trying to use the cross-country evidence to make a solid inference about the effectiveness of aid. But the Bank reports prepared for Monterrey did not present a balanced picture of the research, with appropriate reservations and skepticism, but used it selectively to support an advocacy position. Once again, we emphasize that we do not think that the research was unusually weak relative to the literature. Nor do we challenge the appropriateness of the Bank’s making the best possible case for its policies. But once the evidence is chosen

selectively without supporting argument, and empirical skepticism selectively suspended, the credibility and utility of the Bank's research is threatened.

There is a similar set of issues with the paper "Growth is good for the poor" which is sometimes used to argue that, in the presence of economic growth, explicit anti-poverty measures are redundant. Yet, here too, there are serious questions about whether the conclusion is really supported by the evidence. Their measure of the incomes of the poor (the average per capita income in the bottom fifth of the population) is derived from aggregate national income using either estimates of the share of the bottom quintile from surveys, or from estimates of Gini coefficients of income inequality together with the assumption that incomes are distributed according to the lognormal distribution. The problem is that many of the estimates of the income shares and of the Gini coefficients are quite imprecisely measured and, when the data are uninformative about the true level of inequality, Dollar and Kraay's procedure guarantees that, on average, the incomes of the poor will track average income. If the Gini coefficients were random numbers, the conclusion would be guaranteed. So, in the end, we do not know how much of the result is genuine, and how much is driven by errors in the data.

In this case too, there was a very different view in other Bank research, in this case by Branko Milanovic, who was providing extensive empirical evidence of increasing income and consumption inequalities in the world, and taking a much more jaundiced view of the benefits for the poor of growth and of globalization. Milanovic's results have been criticized by others, and the panel takes no view on the issue, but there is certainly no consensus that his findings are incorrect. Yet once again, the official position of the Bank gave selective prominence to one set of views (for example, in the Monterrey document),

although it is does not appear to the panel that one set of results is any stronger than the other.

Pensions and insurance

Another prominent example of research that is both useful and flawed is the work on pensions and insurance that was reviewed by Peter Diamond (MIT) who read seven flagship volumes on these topics. Diamond finds much to commend in these volumes, which often provide an effective bridge between the theory and empirical literatures, and provide help and guidance to policymakers who have to make sometimes difficult and complex decisions about the reforms of their national pension and insurance schemes. Bank research on pensions has been particularly useful for its focus on providing information about Non-financial Defined Contribution (NDC) schemes. Yet Diamond concludes that “there has been too much advocacy at the cost of more balanced, and so more educational, presentations.” He quotes one of the Bank volumes which complains about “a near-religious war about the virtue of funded versus unfunded provisions, and the merits of defined-benefit versus defined-contribution plans.” But then goes on to observe that

It should be recognized that the Bank economists set (and sustained) the tone for these interactions. Overselling first the value of funded privately-managed individual accounts and then of NDC systems does not serve the Bank’s central role in broadening the understanding of development policy, as stated in the charge for this report. Indeed repeated analytical errors associated with overselling prior views casts doubt on the World Bank pension role.”

The analytical errors referred to are those that would be well understood by a first-year graduate student in economics. Diamond argues that this sort of work, which goes into

country policy work with the weight of World Bank research behind it, should first be evaluated by some peer-review or other evaluative process.

Infrastructure

A different set of concerns was raised by Edward Glaeser (Harvard) in his review of the Bank's research on urban development. The main problem here is that the whole area of urban economics, although vitally important to the Bank, is not thriving outside the Bank. Glaeser writes "The central problem with urban research at the World Bank is that urban studies remains an intellectually challenged field, yet despite the weaknesses of the field, the Bank must remain committed to the area." While the studies that he reviewed were weak, they were "neither unusually good nor unusually bad relative to the standards of urban economics." It is clearly a lot to expect Bank researchers to do good research when no one else is doing it. Yet it is worth noting that in environmental economics, where there is a similar problem in the academic field, our reviewers liked much of the work that they saw.

Poverty mapping

The panel also has concerns about the Bank's work on poverty mapping, another case where there is a publication in a leading journal (a note in *Econometrica*) and where the techniques first presented there are being widely applied in the countries. This research is innovative and addresses an important need. It seeks to supply countries with poverty estimates at a small-area level, for towns, cities, and municipalities, and in some cases, these estimates are used by governments to target financial transfers from the center to

particularly needy areas. For obvious reasons, such estimates are widely appreciated by politicians. The panel was provided an extensive list of countries that are now using poverty maps provided by the Bank's research group.

The evaluation of this work is important, not only on its own account, given its wide use, but because it highlights one of the themes of this report, which is the quality of the measurement and statistical services that the Bank provides to its clients. Poverty reduction is the yardstick by which the Bank seeks to be judged, so that the accurate measurement of poverty underpins all of its activities, just as the appropriate evaluation of projects is at the core of its poverty reduction measures. In both cases, the statistical methodology is key.

The basic idea of poverty mapping is straightforward. Most countries have a recent census which, although it does not collect data on income or expenditures that would permit small area poverty estimates, contains data on a range of other variables that are correlated with poverty, such as education, landholdings, occupation, and demographic structure. Any household survey with income or expenditure data can be used to link the census poverty correlates with actual poverty, allowing calculation of a set of numbers that can be taken back to the census and used to impute poverty estimates from the variables that are included in the census. The household survey provides us with what is effectively a table of poverty rates according to, say, land holdings and the education of the head of household, so that when we go back to the census, we can "look up" any given household in the table, and average over a group of households to get poverty rates for a town or city.

The difficult and contentious issue with this work is the accuracy of these estimates, and indeed whether they are accurate enough to be useful at all. Ideally, the users of the maps, policymakers and statistical offices in the countries that use them, should be able to judge whether the maps are accurate enough for their purposes, some of which, like the allocation of poverty-relief funds, are extremely politically sensitive. To this end, the poverty mapping group at the Bank calculates standard errors for the poverty estimates for each place. However, the panel has questions about whether the Bank's estimates of these standard errors are themselves accurate. This may sound like a technical, not substantive issue, but that is not the case, or at least we must sometimes recognize that what seem like technical issues are central for policy. What we are most concerned about is the possibility that the Bank is making very attractive poverty maps, whose precision is not known, but which come with statements of precision that we suspect may be seriously misleading. In the worst case scenario, a map that may be worthless is presented as one that is extremely precise.

Why is it so difficult to make these maps? And what is the problem of using the census information to provide some idea of local levels of poverty? On the latter first, it is clearly informative to know that a town, city, or small area is particularly well-endowed with an educated population, good infrastructure, and so on. But the leap from there to a poverty measure is clearly a large one. Some places that are not well-endowed do very well (at a global level, think of Singapore), and some places that are well-endowed do very badly (think of Sudan, always tomorrow's "breadbasket" of Africa, and with a well-educated elite). So knowledge of the correlates of poverty is very different from having a good estimate of poverty itself. And given the pervasiveness of the fact

that some places do collectively better or worse than they ought to, there is the potential for a large margin of error. Knowing the size of that margin is crucial if policymakers and others are not to be misled by poverty maps that give a false sense of precision. For reasons that we explain in an Annex to this chapter, we doubt that the methods used by the Bank to assess margins of error are generally adequate. Indeed, we are unable to rule out the possibility that the true margins of error are many multiples of those that are presented.

Unfortunately, we also do not know how to make the appropriate corrections, nor is it even clear that it is possible to do so in principle. And if some way cannot be found to give reliable statements of precision, poverty maps are of very limited usefulness.

The demand for small-area statistics, including small area poverty statistics is a global demand, as strong in rich countries as in poor. So it is worth noting that the US Bureau of the Census has in recent years spent many millions of dollars on the *American Community Survey*, which is aimed at providing for the United States exactly the sort of local information that the Bank is providing to its clients through the poverty maps. Its aim is to provide data for every county, town, and community between the sizes of 20,000 and 65,000 persons. It would be a good idea for the Bank group to exchange views with the Census Bureau statisticians, so that, if poverty mapping can be made to work reliably, the US can have the benefit of it. Or the Census statisticians can explain why they did not follow the poverty mapping route in their own work. If, as we suspect is the case, the method cannot reliably deliver what it claims, a usefully precise poverty map, the Bank should not provide statistical methods to its clients that have been rejected as inadequate by the US statistical service.

Our recommendation is that this work be put on hold until the statistical problems are resolved. It would make sense, for example, to engage a small review panel of statisticians or econometricians with expertise in the area, but certainly including one or more of the small-area statisticians from the U.S. Bureau of the Census.

In the same context, we are also disturbed by the experience of poverty mapping in South Africa, as described in the evaluation by Murray Leibbrandt and Martin Wittenberg (University of Cape Town). In this case, the census was made available to the World Bank team, but neither to the South African Ministry of Finance nor to the advisors who were charged with using the poverty map to allocate funds. This failure of data sharing certainly reflected local problems, and should not be blamed on the Bank team. But as a result, the poverty maps were never locally owned, with the technical work being done only in Washington, and members of the Bank team were not responsive to (informed) technical questions from South Africans working on poverty measurement, including South African academics working with the Ministry of Finance. In the end, the process collapsed, because the South Africans could not integrate their local knowledge and concerns into the mapping and were deeply suspicious of the numbers supplied by the Bank who, in turn, were not able to (or at least did not) clarify the technical questions. Given the very considerable local statistical and econometric expertise available in South Africa, particularly in the Ministry of Finance, we can reasonably suppose that if this process is not locally owned and locally responsive in South Africa, it is even less so elsewhere.

Like the pension work, and the work on the effectiveness of aid, the poverty mapping work is an example of research that was put into practice, with important consequences,

but without an appropriate appreciation of the unresolved technical and theoretical difficulties in the supporting research. The panel also believes that this work illustrates a more general lack of statistical oversight in the Bank, an issue to which we will return.

Civil war

Daron Acemoglu (MIT) praised the Bank's work on civil wars, their causes, and their negative effects on development. He notes that civil war is an extremely important topic but one that is seriously under-researched, so that these volumes, mostly associated with the work of Paul Collier, are particularly to be welcomed. "This may be the most important question for development in sub-Saharan Africa, and perhaps in other parts of the world." He particularly praises the country level detail in these reports. However, Acemoglu also strongly criticized the work for its lack of an appropriate conceptual and empirical framework. As a result, the regression analyses in these studies cannot be used to support the conclusions that they ostensibly reach. As was the case with the very different poverty mapping work, an important and promising topic was marred by poor execution.

Finance and growth

A final example in this section is provided by the Bank's research on finance and growth. According to Marianne Bertrand (University of Chicago), who evaluated much of this work, "there has been over the last 15 years or so a strong interest in documenting cross-country differences in financial development and financial structure, studying the country-level determinants of financial development and financial structure (such as legal

and institutional factors), and assessing the impact of financial development and financial structure on country-level economic outcomes (such as country-level growth). This cross-country research agenda has been quite successful from a publication standpoint, with many papers landing in top finance journals, such as the *Journal of Finance* and the *Journal of Financial Economics*. A non-negligible share of these publications are authored or co-authored by World Bank researchers.” She compliments the Bank researchers for their contribution to constructing many of the financial indicators, and for tackling these questions, which are arguably fundamental for the understanding economic development. However, she criticized the work for its over-reliance on cross-country comparisons and cross-country regressions which, while informative about general patterns, are generally hard to interpret and to use as the basis for policy conclusions. She was surprised that the research had not made a better and more systematic attempt to link the cross-country evidence to country level case studies, in which the Bank surely has a strong comparative advantage.

Jonathan Morduch (New York University), who also read papers in finance, echoed Bertrand’s complaints about the overuse of cross-country regressions, their difficulty of interpretation, and the relative under representation of case studies. He also lamented the recent absence of Bank researchers from discussions about microfinance, a topic that has been largely left to the advocates and to the NGOs, who do not produce the sort of solid and balanced empirical evidence that could have been provided by Bank researchers.

General themes: strengths and weaknesses

The evaluators' comments and the panel's own reading identified a number of general themes that cut across many different studies, and are conveniently gathered together, rather than dealt with study by study, although some have already arisen in the previous two subsections. Again, while the tone of these comments is often critical, the panel is fully aware of the difficulties of meeting all of the targets listed in Chapter 1, and offers its criticisms in what it hopes will be seen as a constructive spirit.

Execution and methods

There is a general feeling that the execution of much of the work falls behind best practice methodology in the profession. The panel understands very well that this is to be expected, and to some extent is desirable; academic methodologies are subject to fads, and it makes no sense for Bank researchers to adopt techniques that are currently being tried for the first time. We also understand that, in a retrospective review such as this, we are looking at studies, some of which were carried out ten years ago, and that would likely be done differently today, even by the same researchers. We should also emphasize that there is no agreement in the profession on what is best-practice empirical methodology in economic development, as indeed will be obvious from a complete reading of the evaluators' comments. There is particular disagreement on empirical methodologies for a central part of the Bank's mission, which is how to evaluate its projects, and how to learn from them. So Bank researchers are admittedly presented with something of an academic minefield, and it is no surprise that they have not always been successful in negotiating it.

It is clear that the Bank is aware of changes that are going on outside, although the adaptation to that awareness is uneven over research projects and researchers. Evaluators felt that a good deal of what they read suffered from the mechanical application of technique, without a sufficiently thoughtful understanding of whether it was appropriate. Of course, a great deal of academic research suffers from the same problem, with many papers in which the ostensible topic is only a pretext for displaying technical expertise. Unfortunately, academic journals often like such papers, and Bank researchers certainly produce them in response to the need to meet publication requirements. But such problems often show up even in high-profile research. For example, many papers appear to think that the attribution of causality is something that can be solved by technical means, through the application of frontier econometric methodology, without understanding that causality can only be addressed by some sort of conceptual originality that identifies circumstances in which causality is running in one way and not the other, and so allows separation of one direction of causality from the other. This problem is particularly apparent in the cross-country work, although there is perhaps less of that now that Bank research has moved away from cross-country approaches to the understanding of economic growth.

Computable general equilibrium techniques

Evaluators were also sometimes critical of the use of what are known as “computable general equilibrium models.” These are essentially large spreadsheets, or simulation models of an economy, and are useful for illustrating possibilities, for demonstrating the sort of thing that might happen if a policy is changed. The term “general equilibrium”

refers to the fact that these models are designed to trace all of the effects of a policy as they spread through the economy, for example as consumers and firms respond to a change in a tax or a tariff, something that is frequently difficult using other methods. The problem, of course, is that probabilities are not the same as possibilities, so that the models are much less useful for predicting what will actually happen, or whether one policy is better than another. As more and more data have become available, and in quantities that could hardly have been imagined a decade or two ago, researchers have become increasingly ambitious in their attempts to provide empirical answers to questions that could previously only have been addressed by simulations. Our evaluators, particularly of the Bank's work on trade, felt that Bank researchers have been slow in moving with this trend.

Project evaluation

Appropriate methodologies for project evaluation are central, not just to Bank research, but to the Bank's objectives in the world, and are currently particularly contentious in the academic literature, and no less so in Bank research. There is a strong recent push for randomized controlled trials (RCTs), and although they are hardly a panacea and cannot be applied to much of the Bank's current portfolio, the panel certainly endorses the view that the Bank should do more of them in cases where they are possible. Bank researchers have been involved in a number of RCTs, although several of those that appeared in the studies that were reviewed were flawed in a number of ways. RCTs are not straightforward to execute; while they make the statistical analysis of outcomes straightforward, the previous econometric expertise must be replaced by very careful

planning and execution in advance. It is clear that the appropriate knowledge was not fully in place in the studies that we saw, but that is to be expected at this stage.

We were more concerned with the presentation of some of the other methods. One example is the technique of matching, where the idea is that those who have been affected by the policy or program are compared with people who are similar in as many respects as possible. This has the disadvantage over RCTs in that it is not possible to control for differences that are not observed, but it is nevertheless an extremely useful methodology when considered as one in an arsenal of methods. Yet several of the papers that were reviewed tended to oversell the matching method, without appropriate acknowledgement of its weaknesses, or even without displaying any good understanding of when it is best to use one method and when another. The Bank has many opportunities to influence how these program evaluations are carried out around the world, and is looked to for leadership; so that it needs to demonstrate a more nuanced and balanced approach in its own work.

Analytical narratives

Several evaluators commented that the work that they liked best, and learned the most from, belonged to a class that we refer to as “thoughtful analytical narratives.” Such research is often supported by (although not dominated by) empirical evidence and econometric results, but it is always deeply informed by economic analysis, which provides a principled framework, and most of all by country knowledge and experience. Local information, and accounts of Bank projects and policies, are areas where Bank researchers have an enormous advantage over other researchers. Indeed, one of their most

important roles is to distill knowledge from operational experience, and a part of that distillation is surely an intelligent account of what happened. Such narratives need to be supported by empirical analysis, but the empirical analysis also requires the narrative support to be appropriately interpreted. These narratives were sometimes case studies, and sometimes discussions of the details of projects and policies, what happened, how they worked, and their immediate effects. As in the rest of the profession, formal empirical analysis has tended to replace analytical narrative in the Bank's work, and the once staple price-theoretical analysis has become less common. Yet it is not clear that this trend is always well suited to the comparative advantage of Bank researchers which lies in the details of country and operational experience.

Use of non-Bank consultants and researchers

Another general issue is the role in Bank research of non-Bank staff. Our evaluators noted that some of the very best work that they read was carried out jointly by Bank consultants, usually in conjunction with Bank researchers, and occasionally by consultants working alone. Yet it was also the case that some of the weakest work also involved Bank consultants. These different outcomes appeared to be predictable in advance from the previous track record of the consultants; weak research was done by people who had previously done weak work, an outcome that suggests a failure of monitoring and management. The use of consultants was uneven and apparently haphazard between Bank researchers and Bank projects, and it appears as if some groups or individual researchers are much more successful than others in selecting consultants, or better at working with good consultants. It seems that weaker researchers tend to select

weaker consultants as in Hirschman's law of squares which posits that first rate researchers make first rate appointments; second rate researchers make fourth rate appointments, and so on.

It is also notable how little of the research was co-authored by researchers from developing countries. We understand that it is often difficult to find first-rate researchers who are prepared to act as counterparts, and know from first-hand experience how pressure to work with local consultants leads to the *pro forma* employment of one of a cadre of local "usual suspects" whose main skill is in being paid to be counterparts. But the benefits of local ownership of research are real enough, and the policy implications of research are more likely to be adopted when the research is thoroughly understood, defended, and propagated by a local researcher. One positive example is cited by Marcel Fafchamps (Oxford) in his evaluation of some of the African research. This is the flagship report *Can Africa claim the 21st century?* notable, not so much for any new research findings, as for the fact that it represented a collaboration between researchers from the Bank and from the best economic research institutions in Africa. That the report's message of growth, trade, and poverty reduction could be jointly endorsed by this wide range of researchers is in sharp contrast to previous disagreements and marks what Fafchamps calls "the beginning of a new era."

Heterogeneity of quality, jumping to conclusions, and self-citation

Nearly all of the evaluations, even those that are positive about some of the work, argue that there is a high fraction of undistinguished work. Daron Acemoglu argues that nothing of what he read would merit publication in a serious general interest journal such

as the *Review of Economics and Statistics*, of which he is an editor. Gordon Hanson (UC San Diego), also an editor, made a similar, if somewhat more favorable judgment about the work that he read in trade, in this case with reference to the *Journal of Development Economics*. Of course, Bank research is concerned with much more than academic publication, so that the failure to meet the standards of the *Review of Economic Statistics* or the *Journal of Development Economics* would only be a serious flaw if *all* Bank research fell into this category. And in fact, the Bibliography for the period lists 11 papers by Bank staff or Bank consultants in the *Review of Economic Statistics*, plus 5 in a special issue on savings; there are 60 papers by Bank staff or consultants in the *Journal of Development Economics*.

But Acemoglu, Hanson, and the other evaluators identified something of a common style for many of these less distinguished papers. They tend to be more academic than policy oriented, the technical execution tends to be weak, and although they nearly always contain policy conclusions, the conclusions are rarely well based on the preceding analysis. We suspect that the relentless pressure to give every paper a policy conclusion, whether or not it actually has one, is largely responsible, though it was not always clear that researchers understood the limitations of their work, let alone communicated it to their readers. As a result, these papers are not making up in policy relevance for what they lack in academic interest; they meet neither of the Bank's criteria. Again, some of this is probably inevitable; the Bank (rightly) expects that researchers will publish papers in academic journals, and the pressure to meet that criterion will surely result in the production of at least some overly academic papers of less than stellar quality. But the panel and the evaluators were surprised by just how much of this work there is, and felt

that better management and evaluation could lead to fewer papers that were neither of academic nor policy interest.

Evaluators also noted that a high proportion of the citations in this group of papers are to other Bank papers, many of them unpublished. In some cases, where groups are almost entirely inward looking, the degree of self-reference rises almost to the level of parody. Again, we must note that there are important areas for the Bank, including such vital areas as poverty measurement, environment, urban studies, and infrastructure that have been “orphaned” by academic researchers, so that there is little good work for Bank researchers to draw on or even to cite. Even so, we suspect that some of the research groups are too inward looking, and have been so for too long.

Missing areas?

Evaluators were asked whether there were important areas in which there was too little Bank work. We have already noted the case of microfinance, and in our interviews with research staff, the absence of work on TRIPS (Trade Related aspects of Intellectual Property rightS) was noted. Both of these seem to be important omissions. Bank research also has some difficulty in the academically orphaned areas of urban studies and infrastructure, although it has done good work on environmental issues in spite of the limited amount of good academic work. The Bank has also scaled back at least some of its traditional work in macroeconomics, particularly on the determinants of growth. However, our evaluators were comfortable with this. Francesco Caselli, in particular, argued that the Bank’s switch towards the investment climate work, with a focus on the

behavior of firms, is very much the right way to go and reflects (and indeed helps lead) similar developments in academic macroeconomic research.

Academic versus policy agendas

Although all but two members of the group of evaluators and panel together are academics, the general feeling was that much of the research was *too* academic, too focused towards the previously existing *academic* agenda, and too directed towards technical rather than pressing policy issues. Less surprisingly, this view is shared by several of the operational staff in the Bank. It is clearly important that researchers in the Bank are seen as meeting high standards of technical competence, and it is hard to see how this can be done without rewarding Bank staff, at least in DECRG, for publishing in academic journals. It is also probably true that academic journals and their reviewers, particularly those not in the very top tier, tend to reward technical, within-paradigm work.

Yet there is a double loss here. Bank researchers are losing the opportunity to tackle policy questions on which they have unparalleled access to data and other information, while the journals are losing the interest that comes from material that addresses new questions and sets new agendas. The best Bank research addresses new questions in new situations in a way that is of wide interest to an academic and general audience, as is well attested by the substantial number of Bank papers published in the leading general interest journals. Many of us who have worked as consultants for the Bank are attracted precisely by its endless supply of new, important, problems that are much more consequential than the latest wrinkle in a well-worked academic literature, and it is clear that this is also a major attraction for the best researchers in the Bank. So the problems lie

not with the best researchers, and the best journals, but with the long tail of undistinguished work that is directed towards, and appears in, the second tier field journals, or in (some of the) conference volumes. While recognizing that not all of this work comes from DEC, we suspect that its amount would be reduced if management relaxed the publication requirement for DEC researchers, and put greater emphasis on other attributes of the work. But such a change would also have costs, and we shall argue below that there are other methods to improve the average quality and relevance of the work and we believe that these other methods should be tried before relaxing the publication requirement.

Dissemination: closing the loop

A related issue is whether, if things work as they should, and good Bank research comes out of operational experience, the results make it back to the country or countries concerned, as opposed to being aimed at academic journals. Academic journals are an important medium for the recording and dissemination of the knowledge that researchers distill from operations, and that distillation is one of their most important tasks. Yet the circle needs to be closed, and the results appropriately disseminated, not least to the originating country.

The World Development Reports

A disclaimer

The World Development Reports, from 1998/99 through to 2006, were read by the panel members, and the comments that follow pool our views. We should note that one of us

(Nora Lustig) was jointly responsible (with Ravi Kanbur) for the *Attacking Poverty* WDR of 2000/01, and both Banerjee and Deaton have provided advice and consultation on several reports. We also note that, as a group, we are far from representative of the intended audience. Some of the material is very familiar to us, so we may give too little credit to the (considerable) expository and communicative value of the reports and, at the same time, we may occasionally be more irritated than others when we perceive faults in what we (think we) know.

Background

The World Development Reports are the most important flagships produced by DEC. They are responsibility of the Chief Economist, who has the opportunity to use them to change the debate on some aspect of development, both inside and outside the Bank. The 1998/99 report on Knowledge and Development covered a topic with which Chief Economist Joseph Stiglitz has long been associated. The 2005 Investment Climate report marked a major new approach to growth introduced by Chief Economist Nicholas Stern, and the 2006 Report on Equity reflected François Bourguignon's commitment to the issue. Other WDRs revisit central issues, such as poverty, or update areas where there has been new thinking inside or outside of the Bank.

Historically, some WDRs have been of lasting importance and influence. Examples are the 1984 WDR on population, which argued that population growth was indeed a problem for development, the 1990 WDR on poverty, which introduced the \$1-a-day poverty measures, and which marked the Bank's recommitment to poverty reduction, and the 1993 WDR on Health, which introduced Disability Adjusted Life Years (DALYs)

and the Global Burden of Disease, and which had such a dramatic effect on Bill Gates.

We understand that there is fierce competition within the Bank to get topics dealt with in a World Development Report.

Recent World Development Reports:

1998/99	Knowledge for Development
1999/00	Entering the 21 st Century
2000/01	Attacking Poverty
2002	Building Institutions for Markets
2003	Sustainable Development in a Dynamic World
2004	Making Services Work for Poor People
2005	A Better Investment Climate for Everyone
2006	Equity and Development

The reports are written by a fulltime staff of around eight people, specially selected for each WDR, and who devote around a full-time year equivalent of their time. Special studies are commissioned by Bank researchers and by outside consultants. There are wide consultations with others in the Bank and, to varying degrees, with outsiders in Washington and around the world. After publication, the report is taken on an international road trip, and extensively promoted by members of the team. The reports are enormously visible around the world, almost certainly more so than any other Bank publication. They are also widely used for college teaching and widely read within the broad development community, in part for information—and the World Bank’s intellectual leadership in the development debate is seen as underpinned by the WDRs—and in part to find out what the Bank is thinking. So they are important vehicles whereby the Chief Economist and other researchers in the Bank influence the development debate,

as well as development policy, not least by influencing their colleagues in the networks and the regions.

Many strengths

The most effective of the WDRs change the debate about development. To do this, they do not necessarily have to be correct, nor to be widely academically accepted, either at the time of writing or later. The 1984 WDR on population offended enough people that the *National Academy of Sciences* produced an outstanding and still classic report in rebuttal. The global burden of disease and the underlying DALYs have been widely challenged in the literature. But both reports would surely be judged as successes.

Outside of the Bank, UNDP's Human Development Report had an important role in helping to broaden the development debate to include health and education even though its lead concept, the Human Development Index, is an arithmetic average of incommensurable objects that was widely condemned by academic commentators and whose deficiencies have only become more apparent with time.

Given that they have broad ambitions, we should be careful not to treat the WDRs as academic monographs, whose main virtue is to summarize the existing literature.

Nevertheless, they often do an outstanding job of doing just that. All of the reports that we read had at least one chapter, and usually several chapters, that provided first-rate reviews that deserve to be very widely read. These reviews are often based on the academic literature, but bring it into a policy focus. A good example is the discussion of the developing academic literature on inequality, institutions, and growth in the WDR on equity and development. The summaries often also make good use of Bank research and

Bank data. For example, there are outstanding summaries of global poverty and inequality in the poverty and equity WDRs, respectively. Some of the specially commissioned work brings together a whole new body of evidence. Perhaps the best known of these is the *Voices of the Poor* study whose main results were incorporated into the report on poverty. Opinions are still divided on the value of that work, but it certainly had a major impact and changed the views of many development practitioners, including the then President of the World Bank. And then there are the boxes, more than a hundred in some reports. These exploit the enormous comparative advantage that the Bank has in drawing lessons from its experience around the world. They are almost always informative and for some, like the cartoons in *The New Yorker*, are the first (and last) things to be read.

The breadth of the scholarship in the recent WDRs is impressive. Lessons are drawn from literatures well outside of economics, including epidemiology, medicine, education, politics, sociology, and anthropology, and people with knowledge of these fields are often brought into the team. In consequence, the WDRs are now much broader than they once were. Health and education have long been seen as central to reducing poverty and the Bank's knowledge and scholarship in these areas has increased over time. But the sensitivity to non-economic issues has increased in other areas too. The 1990 WDR on poverty was about labor-intensive growth and safety nets. By 2000, the discussion had broadened to opportunity, security, and empowerment, and the report has strong literature reviews on such topics as social exclusion and gender.

In all of the WDRs that we read, there are deeply thoughtful discussions of topics that are not always well or widely understood. For example, in the second part of the WDR on

Knowledge and Development, there are chapters that carefully lay out the relevant theory. These do not rely on trying to overwhelm the reader with evidence from cross-country regressions that “prove” particular policy claims. The goal is less to tell us what to do than to help us think through the many issues that arise once one starts thinking hard about these questions, so that we bring a more sophisticated toolkit to the analysis of policy issues. Similarly, the WDR on *Entering the 21st Century* has extremely thoughtful discussions on the role of agglomeration and urbanization in development. The points that it makes have stood up well and the trends that it identifies have, if anything, accelerated. It also contains one of the Bank’s first comprehensive look at the effects of globalization on macroeconomics, finance, and trade. The discussion of trade in this WDR is very good, and anticipates the Bank’s later successful collaborations with the Fund on trade policy issues. Another very interesting and provocative chapter in this WDR (and broadly related to the urbanization theme) is the chapter on decentralization of government. As globalization proceeds, city states suddenly become highly viable, and the case for decentralizing and devolving power to regions strengthens.

The WDR on sustainable development also has many extremely useful discussions. The report takes the view that achieving sustainable development is a matter of constructing the right sort of institutions. It discusses the particular kinds of institutions that are needed and how they might be brought into being, with examples of cases where institutions have been created or reshaped in order to deal with environmental problems as well as cases where the institutions failed to develop. These discussions are not only informative, but usefully suggestive of how policymakers might help design institutions to help deal with environmental issues.

Some weaknesses

The World Development Reports are written by committee, not just by the team members who work together over an extended period and can develop a coherent vision, but by many others who comment and whose views are taken into account. This process makes it difficult to maintain a coherent and focused argument, especially for controversial topics where there is a range of conflicting views, equity and development being a leading example. There is also a tendency to pull political punches so that, for example, large, important countries are rarely criticized, even when the logic of the argument seems to lead in that direction. Issues are seen through the lens of current Bank policies, even when not obviously appropriate. The WDR on *Entering the 21st Century* is burdened with having to mount a sustained defense of the Comprehensive Development Strategy. There is much political correctness, including mindless cheerleading for cultural touchstones such as women, trees, and social capital, as in “women are an important engine of development.”

Trade-offs tend to be eschewed in favor of ubiquitous “win-win” scenarios, so that, for example, growth and environmental improvement are never seen as in conflict, because poverty and pollution are social problems that each mark institutional failure, so that institutional repair can somehow lead to both being dealt with simultaneously. A more equal income distribution is seen as a generally good thing, but there is no discussion of the optimal tax literature that formalizes the necessary trade offs between equity and incentives. More generally, trade-offs between competing goals are downplayed relative to sometimes far-fetched complementarities. While there is

something to be said for such an approach in forging the compromises that are required to make progress in policy formation, it hardly leads to intellectual clarity. The World Development Reports suffer from always trying to make everyone happy.

The committee process also exacerbates the broadening of mission that has characterized the Bank in recent years. While the broadening of the debate can often lead to a more satisfying analysis, it does not always make it easier to think about policy. Safety-nets and labor-intensive growth are areas in which the Bank is likely to be able to offer countries some useful advice. Policies to deal with opportunity, security, and empowerment, or at least for those parts of them that go beyond safety-nets and labor-intensive growth, are more difficult. The 2005 WDR, on the investment climate, is almost a caricature of the view that everything is important. If there are priorities, they are vague and constantly changing, and the report notes that virtually every conceivable aspect of a country's social, political and economic institutions affects its investment climate. There are chapters on Stability and Security, Regulation and Taxation, Finance and Infrastructure, Workers and Labor Markets, and "Confronting Underlying Challenges" (the last including subchapters on restraining rent seeking, establishing credibility, fostering public trust and political legitimacy, .. good institutional fit, etc.) There are 112 boxes that are both interesting and colorful, and cover every region of the world, but their occasional individual excellence only highlights the fact that they do not fit into a consistent argument. Although we certainly acknowledge that the 2005 WDR supported outstanding data gathering efforts, the incoherence of the document itself is unsettling.

The methodological and analytical challenges that we identified in the main body of Bank research occasionally appear in the World Development Reports. As in the general

research, this is particularly problematic when the WDR staff reaches out into new analytical territory. The *Equity and Development WDR* is perhaps the leading example. It contains no stable concept of what it means by inequality; again, we suspect that is a response to the need to try to make everyone happy, even when they have mutually incompatible views. For example, the report never resolves the tension between “equality of outcomes” and “equality of opportunity.” (We suspect that the politically charged nature of topic was also an important factor in this case.) In the last few years, there has been a welcome reduction in the mechanical use of techniques such as instrumental variables, though the validity of studies is sometimes justified by the fact that they used “advanced econometric methods.” We are not always sure that the authors fully understand that technical fixes are no substitute for convincing argument. And when they do, it is the convincing argument that is owed to the readers, not an appeal to the magic powers of “advanced” methods.

Methodology is also a problem when it comes to citing evidence. In the WDR on service delivery, evidence that comes from randomized controlled trials is presented alongside evidence from NGOs whose own propaganda is treated on equal terms. For example, it uses public report cards as one of their central examples of an institutional reform that would help improve the delivery of public goods. A number of places in the report allude to the success of this intervention (for example see page 88), but the supporting evidence comes from a simple before-and-after comparison carried out by the programs sponsor's. This seems to be a very low standard for the evaluation of something that is given so much weight in a WDR. Before-and-after comparisons are always suspect because there can be other things going on at the same time and it is hard to imagine that

this was not true in the fast-changing Bangalore of the last decade, where this study was carried out. And without in any way impugning the integrity of the sponsors of the public reports cards, it is very difficult to be fully objective about the results of your pet project. As with the econometric methods, the appropriate weighing of the convincingness of the evidence is lost. There is much selection of evidence, with obscure, sometimes unpublished, studies with the “right” message given prominence over better and often better-known studies that come to the “wrong” conclusion. While academic studies are by no means immune to such selective citation, there is surely an argument for judicious weighing of evidence and for balance in publications that are so widely read and taught as are the World Development Reports. The Bank devotes an enormous amount of resources to the editing and dissemination of these reports, and it would make sense to devote similar attention to the balance of the evidence.

The World Development Reports are costly and account for around ten percent of DEC’s resources. The need to produce one every year is an enormous tax on Bank research, and the team leaders are often the Bank’s most senior and productive researchers. And in part because of the pressure to include everything, there is a huge overlap from one WDR to another; in particular, there are chapters in the reports on poverty, on sustainable development, on service delivery, and on equity, that are virtually interchangeable. It is hard to believe that it is a good use of such skilled researchers to have them rewrite the same thing in different words year after year.

Some argue that the WDRs typically have a very short shelf-life, though that is clearly not true for a handful of the best known. An appropriate analogy might be with book publishing more generally, where a few successful books on a press’s list provide

the revenue to carry a much more extensive list. Against this it might be argued that the WDRs are supply-driven, and are more used for the internal jockeying for influence than they are used to affect the development debate outside of the Bank. We do not have good measures of internal or external influence, so it is hard to know. Writing about one of the WDRs, the Bank writes:

Since its publication ... the report has been well received globally, and the team has received numerous requests to participate in additional seminars around the developing world to inform policy makers and the public at large of the report's recommendations.

We wonder how much demand there would be for the report's authors' seminar presentations if clients had to pay (say) five percent of their travel costs?

Finally, and surprisingly to us, the quality of editing and presentation is lower than we had anticipated. In one of the more recent reports, there are figures whose axes are unlabeled, and charts with labels that are surely incomprehensible to most readers. Echoing other complaints about the Bank's website, the versions of the World Development Reports before 2003 that are available for download are badly-copied black and white versions of the originals, in which many of the boxes are not legible. This is astonishing for what is, in effect, the Bank's most important and widely read publication.

Chapter 3: Annex 1: Further remarks on poverty mapping

The panel is concerned that the poverty maps constructed by the Bank, and now used in many countries, do not come with adequate warning of their likely inaccuracy. We have no problem with the use of correlates of poverty, derived from the census, to estimate poverty in small areas, using a predictive relationship that is estimated from household survey data in which the correlates can be used to predict the probability of being poor. We have a number of technical concerns about the way in which the estimates are calculated—for example on the robustness of the multi-stage estimation strategy—but our main concern is whether, even in principle, such estimates can be calculated with a useful degree of precision, and whether the standard errors provided by the Bank group accurately assess the precision of the maps.

The statistical issue that most concerns us is that the deviations of local poverty from its predicted value are likely to be correlated across space within the local area. Because labor, commodity, and housing markets tend to be integrated or at least linked by spatial propinquity in a small area such as a city, there is a tendency for everyone in a given city to be either worse off or better off than would be predicted based on a national equation. In consequence, deviations from predictions are likely to be correlated within the small areas for which poverty is being estimated, for example across wards of the same city. In this circumstance, standard errors for the predictions must take into account the spatial linking, if only because the errors of prediction will not average out over neighborhoods in the same city.

If the location effects were independently distributed over city blocks or census “enumeration areas,” (EAs) the small areas used in the census, then averaging over all the EAs in the city would cause errors to cancel one another, and the poverty estimate would become more and more accurate the more EAs that we are able to include. When the Bank group calculates the precision of its estimates, it assumes that the prediction errors in each EA are independent of one another (or at least that is our best reading of the *Econometrica* paper, which is unusually impenetrable, even by its standards.) But the EAs are typically quite small (perhaps 100 to 500 households), and there seems no reason to suppose that, in general, that deviations from prediction will cancel out when averaged over enough EAs in the same small area.

Indeed, if the labor and property markets are integrated at the city level, there will generally be intercluster correlations between the EAs, with all of them tending to be above, or all below the prediction. In such circumstances, adding more and more EAs will never reveal the true poverty rate, because there is something about the city that is just not captured in the correlates. In technical language, *consistent* estimation is impossible. This might not be a problem if standard errors were accurately computed, but the assumption of independence over EAs is used when calculating the standard errors. How big an error this introduces will depend on the number of EAs in the city, and on the intercluster correlation across EAs in the prediction errors. It might be small in some cases, but it is not difficult to construct examples that seem realistic, and where the standard errors are out by a factor of ten or a hundred. Ideally, there would be a method to test whether or not there is a problem. But that would require looking at the correlation between the prediction errors in different EAs in the same city, something that can only

be done in those cases where there are multiple EAs for each area in the household survey. This is certainly worth exploring, but it may not happen often enough to provide diagnostics in every case where they are required, let alone to correct the estimates of precision.

This is a problem that may not have a solution, so that there may be no way, in general, of assessing the precision of poverty maps. This would be consistent with the way that the US deals with its small area statistics.

Chapter 3: Annex 2: Analysis of evaluators' scores

The discussion in the text is based on our reading of the evaluators' reports, on discussions with them, and on our own views. But the evaluators were also asked to provide formal scores of each project that they read and they did so for a number of important dimensions. In this section, we look at these scores, and what they tell us about how Bank research measures up, whether it is stronger in some areas than in others, whether research done in DEC is stronger than research done elsewhere in the Bank, whether flagships are generally better or worse than other research, as well as the strength in various dimensions, such as clarity of exposition, theoretical and empirical analysis, the appropriateness of conclusions, and how firmly they are based on the evidence.

The unit of our analysis is the *project*; each one is sometimes a single output, such as a flagship publication, but more units comprise a set of papers that are the outputs of a single research project. In a few cases, the evaluators provided scores for individual papers, rather than for the project as a whole, and we averaged these to make them comparable with the other scores. For each relevant aspect of the project, evaluators were asked to assign a score of 1 through 5, with 1 meaning unacceptable; 2, below average; 3, average; 4, above average; and 5, superior. The various aspects are listed in Table 1, together with the average scores on each over the (up to) 192 projects that were assessed. The rows of the table are essentially the questions put to the evaluators. We also show the total number of projects graded under each score (not all aspects were relevant for all projects), as well as the mean for all projects, for DEC projects, and for non-DEC projects.

A number of points should be borne in mind in making comparisons between DEC and non-DEC projects. When the sample was drawn, it proved impossible to use the research “codes” under which projects were classified by the research support group, at least for DEC projects. Because funds tend to be moved between projects, the “official” classifications were often not substantively meaningful. In consequence, it was necessary to ask the research managers in DEC to aggregate projects into units that did make sense, from which the sample was then drawn. While we have no evidence one way or the other, it is possible that projects were grouped in a way that would artificially enhance apparent quality, for example by absorbing weak projects into larger, stronger ones. DEC managers were also asked to provide us with their best output, even when it was not included in the sample. Research from outside DEC was not similarly treated. Even so, the panel suspects that the pro-DEC bias is likely to be small, in part because we think that the managers grouped projects in a sensible and substantive way, and in part because they could not easily anticipate which projects the panel and the evaluators would like.

Note also that there is a difference between DEC and non-DEC research in the treatment of “flagships.” There is no completely consistent definition of the term in the Bank, and we used it here to apply to the network and regional publications, usually “glossy” reports, that are primarily aimed at summarizing and communicating the evidence on some important issue. There are no DEC flagships in the analysis here. This is in part because the most important DEC flagship-type publications, the *World Development Reports*, were not sent to our evaluators, but were separately read by the panel. DEC also produces occasional “Policy Research Reports (PRRs),” several of which were included in the sample. However, unlike the regional and network flagships,

the PRRs were often bundled with other research into the parent projects and sent to the evaluators, while a few (five) were treated as stand-alone documents. As a result, we cannot separate out the PRRs in the analysis, and they are treated along with the rest of DEC research. For most of them, this is not only necessary but is appropriate, given that they have a much larger research content than most non-DEC flagships.

Although the numbers in Table 1 are potentially contaminated by the different standards used by different evaluators, they are nevertheless a useful starting point and they provide quantitative support for a number of the points that we have already noted. In particular, the highest of all the scores is for “the importance of the issues addressed,” while the lowest scores are for methods, for the way that conclusions were (or rather, were not) derived from the evidence and for the appropriateness of the recommendations based on the analysis, as well as for the soundness and likely impact of the conclusions for policy. These scores reflect what we have already seen, that great topics are not always addressed with the right methods, and that there is a good deal of jumping to conclusions that are not supported by the evidence, given the deficiencies in method.

Table 1 shows that the scores for DEC and non-DEC projects differ in ways that might be expected. DEC does better on methods, particularly on statistical and econometric methods, and a little better on data, but works on topics that were judged to be less important. Non-DEC projects do a little better on clarity of writing, conclusions, and recommendations, as well as on the appropriateness of recommendations, though even worse than DEC on the relationship between the evidence and the conclusions. The scores for the “overall quality of research” are virtually identical.

To explore the data further, we need to control for the identity of the evaluator. A tabulation of scores for each evaluator gives strong grounds for the suspicion that different evaluators used different standards, although it is always possible that some got lucky and were asked to look at a particularly good bundle of projects. Indeed, this was clearly the case for one or more of the flagship evaluators. Unfortunately, controlling for the evaluator is not an ideal procedure, because it controls for more than we want. For example, the evaluators who looked at the transition and “Doing business” flagships did not score other projects. So that once we remove means for each evaluator, we lose any influence that these scores have, for example, on the differential score of DEC and non-DEC projects., or between flagships and non-flagships Nevertheless, it would clearly also be a problem if we were to ignore the cross-evaluator differences in average scores. For obvious reasons, projects were not randomly allocated to evaluators, so it is always going to be difficult to separate out differences in evaluator standards from differences across areas in the quality of the work.

Table 2 shows the scores attached to being a DEC project or being a flagship project for the same topics as in Table 1. These are obtained by running separate regressions for each type of score (i.e. one regression for clarity of exposition) on a set of evaluator dummies, a dummy for whether or not the project was a DEC project, and a dummy for whether or not it was a flagship. Each row of the table shows the results of one regression, and shows only the coefficients on the DEC dummy and on the flagship dummy, as well as the absolute *t*-values; coefficients with *t*-values of 2 or more are shown in bold. The omitted, or baseline, project is a non-DEC, non-flagship project, so that the numbers in the table are the scores relative to that class of project. Numbers

should be read across the rows, and cannot be compared down the columns, so this table does not tell us the strengths and weaknesses across aspects, only across flagships versus non-flagships and DEC versus non-DEC. (We will return to the cross-aspect evaluation below)

Of the 192 projects, 135 came from DEC. Of the 57 projects that are not from DEC, there are 26 flagships and 31 non-flagships.

All but three of the numbers in Table 2 are positive, and all of those that are statistically significant are positive, which says that our evaluators generally rated DEC projects and flagships higher than the omitted category of non-DEC, non-flagship projects. The comparison of DEC versus non-DEC flagships (remember we have no DEC flagships) is not consistent across aspects, although DEC projects are clearly stronger in statistical and econometric methodology, and in data handling. This is consistent with Table 1. On the score for “overall quality of research,” DEC research projects do a little better than non-DEC flagships, although the difference is not statistically significant. (Both do significantly better than non-DEC non-flagships.) There are no cases where the flagships do *significantly* better than regular DEC research.

We have also considered whether the evaluator scores provide useful information on research areas, such as macroeconomics and growth versus infrastructure, but this turns out not to be possible in any useful way. The evaluators were chosen according to their areas of expertise, so that it is difficult to sort out differences in grading standards from differences in area quality, in spite of the fact that some evaluators saw more than one area, and some areas were covered by more than one evaluator. For some areas, there was only one evaluator, who read only in that area, so that it is impossible to separate the

evaluator effect from the area effect. Given our inability to compare *all* areas, we chose not to try to grade some. Although we are confident in our earlier assessments of area weaknesses and strengths based on the comments and discussions, the lack of a precise quantitative characterization should be born in mind.

Our final analysis concerns the strengths and weaknesses of Bank research on the various different criteria put to the evaluators. We can do this by assuming that the each evaluator's generosity (or lack of it) is constant across areas, so that it is a pure "personal" effect rather than a person interacted with question effect, so that each evaluator's score on a particular question is the sum of his or her personal effect plus a question effect. We obtain these estimates by running one single regression of all scores on all questions by all evaluators against a set of dummy variables, one for each enumerator, one for each question type, as well as dummies for whether the project contains a flagship, or whether it was done in DEC.

The results are shown in the first two columns of Table 3. As before, both DEC and flagships do better than non-DEC and non-flagships, with regular DEC research doing better than flagships by a small and insignificant margin. Over the various characteristics of research, here graded relative to overall quality, Bank research does particularly well on the importance of the issues, the clarity with which they are formulated, and getting its work out in the appropriate format. It does less well on its use of statistical and econometric methods, on empirical analysis, on providing a sound basis for policy, on influencing governments and the development community. We attempted to disaggregate these effects by DEC and non-DEC, differences that are likely present given the previous analysis. However, when all the necessary interactions are included, the results are too

imprecise to be helpful. Even so, the numbers in Table 3 should be interpreted as averages over DEC, non-DEC, and flagships or not.

The final two columns of Table 3 repeat the analysis with the inclusion of indicators of the size of the project, where we distinguish small projects with a budget of up to \$75,000 and large ones, with a budget of at least \$400,000. There is an abbreviated approval procedure for these small projects, so it is possible that they are assessed differently. We do not have budget information for the flagships, nor for a number of DEC projects, including the “best-in-show” outputs that were added later, and the PRRs. In consequence, we confine the analysis to the 140 remaining projects, which excludes all of the flagships. The results for these projects, over the different aspects of Bank research, are very similar to the results for the full sample in the first two columns. The margin of DEC over non-DEC is still substantial, and statistically significant. The small budget projects attract a significantly lower score than the others. It is not entirely clear what to make of this finding. The scores are for the project as a whole, and tell us nothing about value for money; the smaller projects may be graded less well simply because they are smaller, and produce fewer outputs. But note that there is no premium for the very large projects, whose scores are also lower, though not significantly so.

Summary

The quantitative conclusions in this section are consistent with our previous analysis based on the evaluator reports and on discussions with them. Bank research is strong on importance, and less strong on execution, and on drawing appropriate conclusions from the work. The weakest kind of research is “regular,” i.e. non flagship, research outside of

DEC. One surprise is that the evaluators liked the flagships which were ranked as highly as regular DEC research, although there was a substantial gap between flagships and DEC together, on the one hand, and regular non-DEC research on the other.

Table 1: Averages scores by aspects of research

Aspect of research	# projects	All	DEC	Non-DEC
<i>Topics</i>				
Importance of Issues	180	4.22	4.16	4.39
Clarity of project focus	163	3.85	3.81	3.98
<i>Analysis</i>				
Theoretical framework	159	3.36	3.43	3.15
Empirical application	151	3.30	3.40	2.91
Statistical & econometric methods	142	3.15	3.22	2.85
Use of existing knowledge	151	3.65	3.66	3.64
<i>Data</i>				
Awareness of other data	134	3.77	3.78	3.70
Compilation, cleaning, etc.	122	3.71	3.72	3.67
Survey design & sampling	79	3.64	3.70	3.37
<i>Quality of output</i>				
Clarity & organization of writing	176	3.86	3.85	3.86
Clarity of conclusions & recommendations	172	3.70	3.67	3.76
Conclusions based on evidence?	174	3.43	3.48	3.33
Appropriateness of recommendations	161	3.44	3.40	3.55
Appropriateness of output form	173	3.73	3.71	3.80
<i>Extent to which research:</i>				
Increases knowledge & understanding	176	3.68	3.70	3.63
Provides a sound basis for policy	175	3.37	3.33	3.47
<i>Actual or likely impact of research on:</i>				
Government policy	176	3.37	3.30	3.53
Future analysis	178	3.52	3.58	3.37
Development community in general	178	3.27	3.29	3.24
<i>Overall</i>				
Overall Quality of Research	192	3.56	3.58	3.51

Table 2: DEC v non-DEC and flagships versus non-flagships by aspects of research
(Improvement in average score over a non-DEC non-flagship research project)

Aspect of research	DEC	t-value	FLAGSHIP	t-value
<i>Topics</i>				
Importance of Issues	0.19	1.1	0.23	0.9
Clarity of project focus	0.30	1.4	0.90	1.6
<i>Analysis</i>				
Theoretical framework	0.30	1.6	0.30	0.5
Empirical application	0.52	3.0	-0.55	1.0
Statistical & econometric methods	0.51	2.5	-0.35	0.6
Use of existing knowledge	0.30	1.7	0.46	1.0
<i>Data</i>				
Awareness of other data	0.32	1.9	0.00	0.0
Compilation, cleaning, etc.	0.25	1.2	-0.91	1.4
Survey design & sampling	0.46	1.8	1.46	1.4
<i>Quality of output</i>				
Clarity & organization of writing	0.69	4.1	0.72	2.7
Clarity of conclusions & recommendations	0.53	2.7	0.66	2.1
Conclusions based on evidence?	0.38	2.1	-0.23	0.8
Appropriateness of recommendations	0.31	1.7	0.42	1.4
Appropriateness of output form	0.51	3.1	0.77	2.7
<i>Extent to which research:</i>				
Increases knowledge & understanding	0.60	3.0	0.54	1.6
Provides a sound basis for policy	0.45	2.1	0.70	2.0
<i>Actual or likely impact of research on:</i>				
Government policy	0.14	0.7	0.15	0.4
Future analysis	0.71	3.5	0.20	0.6
Development community in general	0.61	3.3	0.48	1.5
<i>Overall</i>				
Overall Quality of Research	0.50	2.8	0.35	1.2

Note: Unlike Table 1, comparisons should be made only within the same row, and not down the same column. Within each row, the number shown is the score relative to a non-DEC, non-flagship project. Each row represents a regression, which also includes dummies for each evaluator.

Table 3. Strengths and weaknesses of research
(Relative to overall quality of the research)

Aspect of research	Score	t-value	Score	t-value
<i>Topics</i>				
Importance of Issues	0.66	7.6	0.64	6.5
Clarity of project focus	0.33	3.7	0.23	2.3
<i>Analysis</i>				
Theoretical framework	-0.17	1.9	-0.09	0.9
Empirical application	-0.22	2.4	-0.14	1.4
Statistical & econometric methods	-0.38	4.1	-0.31	3.1
Use of existing knowledge	0.09	1.0	0.13	1.2
<i>Data</i>				
Awareness of other data	0.22	2.4	0.23	2.1
Compilation, cleaning, etc.	0.17	1.8	0.15	1.4
Survey design & sampling	0.12	1.1	0.17	1.4
<i>Quality of output</i>				
Clarity & organization of writing	0.29	3.4	0.21	2.1
Clarity of conclusions & recommendations	0.13	1.4	0.04	0.4
Conclusions based on evidence?	-0.14	1.6	-0.08	0.8
Appropriateness of recommendations	-0.10	1.2	-0.11	1.1
Appropriateness of output form	0.18	2.0	0.06	0.6
<i>Extent to which research:</i>				
Increases knowledge & understanding	0.12	1.3	0.13	1.3
Provides a sound basis for policy	-0.20	2.3	-0.24	2.4
<i>Actual or likely impact of research on:</i>				
Government policy	-0.20	2.3	-0.20	2.0
Future analysis	-0.05	0.5	-0.01	0.1
Development community in general	-0.29	3.4	-0.33	3.3
<i>Overall</i>				
Overall Quality of Research	BASE			
DEC	0.44	10.2	0.30	6.4
Flagship	0.34	4.4	--	
Budget <= \$75K	--		-0.18	3.2
Budget >= \$400K	--		-0.04	1.1

Note: The coefficients come from a single regression that also includes dummies for each evaluator. The last two columns contain only 140 observations for which we have budget information, and excludes flagships as well as PRRs.

Chapter 4. Evaluator comments by area

The evaluators were selected to cover nine general areas, or fields, and were asked, in addition to their reviews of each project, to comment on the general quality of Bank research in their area of expertise. The areas were (1) macroeconomics and growth, including the investment climate work, (2) fiscal policy, public sector management, and governance, (3) trade and international economics, (4) poverty and social welfare, (5) human development (health, education, population, employment), (6) finance and private sector development, (7) agriculture and rural development, (8) infrastructure and urban development, and (9) environment. In most of these areas, we had two evaluators, and to this team, we added individual scholars to look at a selection of flagship reports, on pensions and insurance, on poverty and corruption during the transition in Eastern Europe, and on the Doing Business surveys. Other flagships were assigned to the field evaluators. Members of the panel read the World Development Reports from 1998. In this section, we provide a summary of the views, area by area. The list of projects and flagships that were reviewed is included in an Annex to the report.

Macroeconomics and growth

Francesco Caselli usefully divides the work in this area of research into three classes of work, (1) data collection, (2) formal theoretical and empirical work, akin to academic research, at least in methods, and (3) informal policy discussions, using theory, empirics, and case studies. The Bank is a world leader in (1) and (3), areas where it has an absolute advantage. The investment climate surveys are only the latest in a series of data

collection exercises, and will undoubtedly have a first-order impact on research and thinking. An even more recent cross-country data set on the value of natural resources is also “superb work” that “is already attracting quite a bit of attention.” The Bank’s work under (3) is the best in the world, clearly superior to similar material from UNDP, and while the IMF is a leader in its own area, it is not a development agency. The Bank’s World Development Reports play a crucial role in both educating and informing both the academic and policy communities. As evidence of this, Caselli cites the fact that it was the 1993 World Development Report on global health that provided Bill Gates with the “moment of truth” that inspired him to work towards the health of the world’s poor. Perhaps the only complaint against this work is that it “sometimes conveys an inflated sense of the solidity of our knowledge in certain areas, and it consequently offers peremptory policy advice that in some cases turns out to have been unwarranted.”

It is the work in category (2) that is most difficult, and whose quality shows enormous heterogeneity including a long tail of undistinguished research. There is work here that is as good as the work produced in the top academic departments, there is much work that is solid and informative, and there is work of poor quality. The worst of the last “is surprisingly bad: poorly written, poorly motivated, and poorly executed.”

Daron Acemoglu, who admittedly uses a very high standard of comparison, finds that almost all that he read was well below academic best practice. The work was typically well-written, usually aware of the relevant literature, and occasionally addressed important and neglected issues, particularly in the books and flagships which were clearly doing a good job of dissemination. Much of the work was flawed by “serious deficiencies in terms of empirical work or conceptual framework.” Like several of our evaluators,

Acemoglu feels that there are many talented researchers in the Bank, but that the work does not always reflect those talents nor use them to the best purpose.

Fiscal policy, public sector management, and governance

The projects that were collected together under this head turned out to be quite diverse, so that the evaluator, Timothy Besley (LSE), offered general reflections on the Bank's work, rather than an evaluation of this specific area. His general evaluation is very similar to that of Francesco Caselli summarized in the previous subsection.

Andrew Foster (Brown), who received a heterogeneous packet of work, looked at a number of papers on village governance. Of these, he wrote: "This work involves detailed analysis of micro-level survey data. This work seems to be driven by important theoretical and policy questions and gives appropriate attention where possible to high standards of inference. The best of the work may be appropriate for first quality academic journals. It also shows a potential comparative advantage of the Bank in terms of implementing large-scale and timely surveys on important policy related issues."

Trade and international economics

As part of his overall evaluation of the area, Sebastian Edwards (UCLA) looked back over the work that the Bank has done in trade and international and notes that, "historically, the Bank has had a very active, vibrant and influential research program on international trade and trade policy." While he commends much of the recent work, and argues that it compares well with academic applied research on the topic, he suspects that it has not been as influential in recent years as once was the case.

Along with almost all of the evaluators, Edwards, together with the other trade evaluators, Gordon Hanson and Nina Pavcnik (Dartmouth), commend the Bank for its recent work on data collection. They identify projects that have generated data on the tariff equivalents of non-tariff barriers, on stocks of emigrants by receiving and transmitting countries, on dispute settlements by the WTO, as well as the provision of software for making better use of trade data from the UN. Yet the evaluators are concerned that these data sets are not all easily accessible by other researchers, and the lack of accessibility is likely to limit their value for generating knowledge about trade, and the software for the UN data is only useful for those who have subscriptions to the original data. Hanson and Pavcnik also argue that the Bank has not done enough to compile comprehensive data on trade costs. They argue that information on “how industries, regions, firms, and households respond to changes in trade barriers” is fundamental to Bank analysis of trade reform.

Hanson and Pavcnik write that “The Bank has succeeded in extending the trade literature to address topics that are important for development policy but that have been neglected by academic research. While this work is not always at the forefront of the literature in terms of technique, it does help guide research toward areas where the social return is high. The Bank has also succeeded in producing excellent syntheses of what academic research has to say about trade policy, which also appear to have a high social return.” On the other side, they argue that “there are also areas where we feel Bank research has been deficient. These include variability in the quality of research (associated in part with the venues chosen as outlets for work), the low production of public research goods, the exclusion of several trade topics important for development

policy, and a tendency to emphasize computable general equilibrium modeling over econometric analysis in empirical research.” The Bank’s work “on product standards, the WTO, international migration, Arab economic integration, and intellectual property rights has helped close gaps in the literature. Underrepresented in Bank projects are papers on trade and growth, trade and institutions, and multinational firms.” Nor has Bank work sufficiently addressed the effects of trade on poverty.

Hanson and Pavcnik argue that “Bank empirical research tends to be dominated by the analysis of ex ante policy changes . . . and the use of computable general equilibrium models. While there are good reasons for this approach, the resulting research portfolio is unbalanced. Missing in Bank research is sufficient attention to analysis of ex post policy changes and the use of modern econometric techniques.” Computable general equilibrium models allow researchers to simulate the economy-wide effects of a policy change, such as a tariff reduction, or the removal of a quota. But as data have become more plentiful, and econometric methods have improved, it is increasingly possible to check the simulations, or even supplant them, with analysis based on the actual data.

Poverty and social welfare

Esther Duflo, in her evaluation, commends the poverty group for its enormous contribution to the measurement of global poverty, and its development of the dollar-a-day counts. She also notes the importance of the agenda on delivering services to the poor, which was the subject of the World Development Report on service delivery which, in turn, generated a number of important background studies. This work helped change the agenda from “how much money is needed?” to “how should it be delivered?” On

other topics, she found the papers on culture to be interesting, but questions the relevance of this agenda to the mission of the Bank. More generally, she felt that this work was less useful than it might have been, in part because of the academic nature of many of the studies, and in part because so much of it seemed to be so much driven by background work for World Development Reports. The academic-style work that appeared in first tier field journals was invariably of interest, but the same could not be said of the substantial amount of work that fell below that standard, even if it was published in lower tier journals

On the issue of self-reference, she wrote that “Some of the papers that I was asked to evaluate seem to be a little trapped in their own world. Cross-references to a very small number of authors in their groups (and not only in the Bank) are very frequent. I don’t think it reflects a lack of knowledge of the outside research, but more likely a tendency to be inward looking. But then, may be this is a general bias and is not specific to the Bank. Many of the papers display good knowledge of the countries.” She also praised the poverty monitoring website, while noting that parts of it (such as the location of the information on the consumption purchasing power parity exchange rates) are very hard to find without expert guidance. Like all of our evaluators, Duflo singled out data collection as particularly useful, although there is wide variability in data usefulness, with some extensions of the LSMS having very high value, while other data appeared to be collected for no particular purpose.

In summary, Duflo writes:

“Generally, the World Bank research is at its best when it does what no-one else has either the incentive or the means to do: assemble large quantities of data (Pro-poor growth)—investing tremendous effort in developing a new data collection tool that will allow to collect comparable data on new issues (Cost of Mental Health—

Das and Hammer); use fine details of a program details to design a credible identification for a project's impact (Some of "poor area" papers); use its leverage with the member countries to allow for randomizing either program placement or the details of the program rules (Reaching the poor: Cambodia—Fighting corruption in KDP).

It is at its worst when it follows some fad either in academia or from within the Bank and comes up with isolated research projects that have little scientific value and no external benefits; or when it collects expensive data sets to estimate program impacts even when it is really not possible to do so."

The second evaluation of the poverty work was prepared jointly by Murray Leibbrandt and Martin Wittenberg. They note that "The Bank has made an enormous contribution in this area, ranging from improving the quality of the data available, to improving capacity in their use, to innovative analyses and careful theoretical work." They note that Bank researchers Martin Ravallion and Branco Milanovic have made major contributions to the measurement and understanding of global poverty and inequality. Like several other evaluators, they note the great value of the LSMS, and commend the Bank's three volume set that documents the lessons that have been learned from the LSMS experience. They also question whether the Bank has been as successful in strengthening local statistical offices, presumably with South Africa foremost in mind, as it has been in collected primary data. They also document some of the problems with the poverty maps to which we have already referred.

They note that the Bank has a track record of compiling data on poverty and inequality and making them readily available on the Bank's website, the most famous example being the Deininger and Squire data on inequality, which is a compendium of secondary data, but also including the data on the Povcal (poverty monitoring) website, where most of the data are derived by Bank staff from primary data or from specially commissioned tabulations. On these, they question whether the underlying data are really

up to the task, particularly when it comes to analyzing changes over time, which is what people are ultimately more interested in than measuring levels. Deficiencies in the inequality data for OCED countries have previously been documented in a well-known paper by Atkinson and Brandolini, but they note that the available South African data present similar problems (as presumably do those for other countries). Without a serious understanding of the way the South African data were collected, and of the structure of the various surveys, all of which were collected in different ways, users of the Bank's compilations would be seriously misled, and the documentation on the websites is not sufficient to allow users to understand the problems. The panel suspects that South Africa is by no means an outlier in this respect.

Leibbrandt and Wittenberg note that the public availability of the LSMS survey data has contributed a great deal to capacity building in South Africa, by generating a demand for econometric training in order to provide local analyses, which in turn fed back into survey design and improvement.

They conclude their overall evaluation with the remarks:

“On the whole we feel that the Bank's research has been of high quality. Indeed a tricky theme that we have highlighted above is the trade-off that the Bank faces between cutting edge innovation in measurement and evaluation techniques versus the harder slog of convincing and empowering developing countries to found their policy making on appropriate data and technical work.”

Human development (health, education, population, employment)

Joshua Angrist (MIT), who was one of our most negative evaluators, found the work that he reviewed to be “variable, running from the best policy-related scientific research I have seen in modern empirical development to reasonably good studies of modest policy

relevance, to studies that were neither very good nor very relevant.” He identifies the lowest quality projects as “purely descriptive studies and impact evaluations without a transparent and compelling identification strategy.” He argues for more randomized controlled trials, particularly in work connected with education, to which such methods are well-suited.

Nancy Birdsall praised Bank researchers writing that the “Bank has made a serious and substantial contribution in health research – especially economics of health, where its staffing, access to micro data covering health and economic characteristics of households for example, has been ably exploited. The Bank research on health systems seems especially important.” Among the projects that she read, she also commended the work on nutrition, and a flagship report on gender. As did almost all reviewers, Birdsall praised the Bank’s work on data collection, particularly the LSMS surveys, but noted that there are still a number of serious issues to be addressed. Data are, or ought to be, a public good, and the Bank often does too little to persuade countries to permit wide access to data collected with Bank funding. As a result, non-Bank researchers get access depending on who they know. She writes, “There has been progress. . . but from 0 to 3 on a ten-point scale, and mostly as a function of individuals not as a function of structural changes.” She argues for more panel data on households, and that the Bank should put money into collecting panel data now, just as when they started the LSMS twenty years ago.

More generally, she worries whether “the skills and experience of Bank research staff are fully exploited in country-specific work. Research staff have incentives to maintain their standing in their fields – and there are fewer risks in doing so with existing data,

using sophisticated tools, as opposed to digging in and understanding countries and their political and institutional characteristics.” She doubts that there is anything like enough dissemination of results, and like several of the evaluators, makes a plea for more randomized controlled trials.

Sebastian Galiani noted the heterogeneity in the projects that he read, all of which were empirical. Some were thoughtful, and a number of them were published in good journals. Much of this work is concerned with evaluation in one form or another, and frequently made use of panel data to compare changes in outcomes, while others relied on cross-section data. A few cases exploit natural experiments of one kind or another. There was one case of a randomized control trial, but it had not been correctly implemented. Much of the work was not innovative, rather replicating earlier studies in a new context; as we have noted, this may well be exactly the sort of research that the Bank ought to be doing.

Finance and private sector development

Marianne Bertrand noted that much of the Bank’s work in finance and development has followed a well-defined program of documenting cross-country differences in financial development and financial structure, looking at the determinants of those differences, and then studying the effects of the differences on economic growth and a range of other outcomes. The questions are of fundamental importance to policy, and to the understanding of economic development. Bank researchers have been leaders in this area, not least in the construction and dissemination of many of the financial indicators. Moreover, and unlike much of the rest of the literature, Bank research has often linked

the indicators to firm level data. This agenda has been quite successful in the academic literature, generating publications in some of the top finance and economics journals.

Bertrand goes on to write:

“While (and maybe because) fundamental, these questions are also extremely difficult to answer convincingly. In particular, the cross-country approach that is adopted in much of the research I have reviewed suffers from serious limitations. While this research approach has established clear correlation patterns between many of the key variables of interest, the policy takeaways of this research are often quite limited due to obvious interpretational issues. Also, this research approach is often too “black-boxy” to provide practical guidelines for those in charge of policy design and implementation.

While I am certainly not advocating abandoning the cross-country research methodology, I was nevertheless surprised by how prevalent this research methodology was in the various projects I reviewed. In particular, I found detailed case studies, where one can delve deeper into the specific experiences of a given country (or a given financial institution within a country), remarkably scarce. My prior going into this evaluation is that Bank researchers had a strong comparative advantage in such case studies compared to researchers at academic institutions, not only given the huge amount of field experience within the Bank but also given the many contacts the Bank has with financial institutions and financial agencies around the world. I was surprised not to see this comparative advantage more strongly reflected in the Bank research.”

While she read some “country case studies that managed to successfully combine a deep contextual knowledge with corroborating quantitative information; in particular, I would highlight here some of the research on the political economy of bank privatization,” she wondered if the incentives for Bank staff to publish in academic journals, which would typically not publish such material, were getting in the way of its production.

Returning to data, Bertrand felt that the Bank had been quite successful in its data collection efforts. The updating of indicators that is currently underway will generate panel data that will help with some of the interpretational issues. She is also enthusiastic about attempts to collect data that help get inside the black-box, for example on micro-indicators of the “reach” of the financial system, or extensions to the credit modules in LSMS surveys. She also noted that some successful pilot programs, for example one that

collected information on bribes in Uganda, and which generated some excellent work, do not appear to have been followed up or replicated elsewhere.

Jonathan Morduch's evaluation is consistent with and complementary to Bertrand's. He writes that "In the main, the research is thoughtful, imaginative, and shows impressive initiative. Many of the working papers are destined for quality peer-reviewed academic journals or are already published there. Many studies rise to a high level of technical sophistication and empirical firepower. Few seem overtly driven by ideology, and a substantial minority question existing or past World Bank policy." But he then goes on to qualify his praise with "At the same time, taken as a group, the projects hold too little immediate relevance for policymakers. Many projects remain at too high a level of aggregation to speak to country-specific debates. Even the country-specific work is nearly entirely divorced from the political concerns of implementers. As such, conclusions are apt to be only modestly useful—and sometimes misleading—for those who most need the results."

Morduch goes on to echo Bertrand's concern with the usefulness of cross-country regressions, particularly for policy work. He also wonders why there is not more reflection of country operational work in these studies, and why there is not more theoretical development. He repeats the almost universal concern of our reviewers that papers aimed at the academic journals tend to have limited usefulness for policies—in many ways as much a complaint about academic journals, as about Bank research—and bemoans the result that papers tend to follow the academic agenda, rather than setting their own. As we have already noted, Morduch feels that the Bank should have been more of a presence in discussions of microfinance, and suggests that more work needs to

be done on what makes financial policies what they are, or the political economy of finance.

Agriculture and rural development

Chris Udry (Yale) argues that the Bank has made important contributions to his main area of expertise—the microeconomics of economic development in Africa—but that it has also missed a number of important opportunities. On the positive side, the availability of LSMS surveys have made an enormous difference to research on Africa, and has generated institution building, for example in the Ghanaian statistical service which is currently engaged in collecting the fifth round of the Ghanaian Living Standards Survey. There has also been important Bank work on HIV/AIDS and health more generally, on civil war, on gender aspects of development, and on various aspects of household behavior.

The missed opportunities are first, the lack of a sustained attention to long-term, systematic data collection. “The Bank is in a unique position that could be leveraged to make extraordinary progress. It has the institutional stability and established global presence to support countries to achieve monitoring capabilities akin to those of the NSSO in India. Data collected on a broad range of activities by households and firms over long periods of time (either in the form of panels or repeated cross sections) that permit surprising and *a priori* unpredictable connections to be drawn are particularly needed.” To which he might have added the ability to make reliable statements about poverty trends in African countries. He writes that “Too often, however, data collection

efforts are hurried, fitful, abandoned, hidden, too narrow, and casual.” Nor is there any systematic policy about making data available to researchers outside of the Bank⁹.

The second missed opportunity is

“the apparent separation between ‘operations’ and ‘research’. I cannot comment on the use of Bank research in ongoing operations. However, it appears that Bank programs and projects offer unexploited research opportunities. Most obviously, this could come in the form of insights to be gained from researchers participating in the design of certain projects. I am confident that several of the other reviewers will be suggesting substantial increases in the use of randomized design in Bank projects, and I am fully supportive. More generally, consultation and collaboration with researchers during project design, coupled with the opportunity for appropriate data collection could open up broad new insights into development processes.”

Udry lists a number of important policy issues for Africa that have not been fully covered by Bank research: (1) program evaluation, for example of service delivery and of microfinance schemes, (2) the effect of institutional innovations, for example the reform of land tenure, (3) infrastructure, (4) firm dynamics, and (5) non-farm enterprises and rural diversification.

Marcel Fafchamps echoes Udry’s compliments and concerns about pluses and minuses of the Bank’s data activities. He rightly points out the difficulties that Bank researchers face in having to satisfy two not always compatible goals, writing research that can be published in academic journals, and writing research that will support and improve the quality of operations. His sense is that the research department is doing pretty much as good a job as could be expected, given the circumstances. Justin Lin is broadly complimentary about nearly all of the work that he read, although there are a couple of projects that reach conclusions that he is unhappy with. Overall, he ranks the work that he read as of “exceptional quality.”

⁹ We understand that the Bank has recently instituted such a policy, though we do not know any details. We look forward to the policy being widely promulgated to the academic and policy community worldwide.

Infrastructure and urban development

Edward Glaeser was the evaluator primarily responsible for looking at the work on urban development, and we have already quoted him on the weakness of the field in general and its importance to the Bank. He goes on to say:

“After all, the developing world is increasingly urban, and development is often so closely correlated with urbanization that it is impossible to think about growth without also confronting cities. Moreover, the Bank’s work on infrastructure often brings it directly in contact with urban policy. Cities usually need more transportation, sewage, and water infrastructure than rural areas. This makes the intellectual difficulties of my own field all the more costly for the Bank.”

The studies that he reviewed were neither unusually good nor unusually bad by the standards of the literature in urban economics. “All of these projects are reasonable, but none of them represent large improvements to our understanding of these topics. Even more problematically from the Bank’s perspective, none of them really changes our view of appropriate urban policy. This does not suggest that the bank’s urban research is unusually bad, but rather in line with the bulk of urban research worldwide.” He found the flagship reports stronger than the research projects, although not without their own flaws. He thinks that, like other projects, urban projects could almost certainly benefit from more randomized controlled trials, but he warns that, given the nature of urban phenomena, where spatial equilibria are of the essence, RCTs will at best be difficult to design, and in many cases will be impossible.

Michael Kremer (Harvard) was the evaluator responsible for the work on infrastructure, and his comments are similar to those of Glaeser. He writes

“Infrastructure is a critical issue in development and one that is very under-researched, both within the Bank and outside the Bank. The research I reviewed included a number of nice pieces such as the Nepal works on specialization and the spatial division of

labor, and on isolation, welfare, and rivalry. The collection of cross country data on telecommunications and electricity regulations was promising. And although this was not part of the review papers, the recent work by Galiani, Gertler, and Schargrodsky on privatization of water infrastructure in Argentina is very good and important. But, based on the research I was sent, I cannot say as a whole that the Bank has really made the contribution on infrastructure it should have. I sometimes had the sense that the people putting together the set of papers had to stretch to come up with enough papers to fit in the category. Thus, for example, two of the documents sent to me were not really research papers at all but rather documents related to proposed bank projects, for example, about the rehabilitation work that would be necessary for a particular power plant to come into line with EU regulations. This may have been a very useful document but I don't think it constitutes research.”

He also noted that most of the work that he read was related to policy only indirectly.

One of the research projects that he liked the best collected comparable international data on telecommunications and electricity regulations, and argues that more such data sets would stimulate research in this under-researched area. He also argues that RCTs are likely to be possible, if not for all infrastructure projects, at least for a good number.

Environment

Geoffrey Heal (Columbia) provides a very positive overall evaluation of the Bank's work on the environment, based not only on the work that he read, but on his previous knowledge. “The research addresses important current issues, is based on a sound understanding of the existing literature, and shows great professional competence. I have no substantial criticisms.” The papers that he reviewed included a mix of work that had been published in the leading area journals, as well as projects that were not research as such, such as on carbon finance and on baselines for Joint Implementation and the Clean Development Mechanism, were socially valuable and were well-executed given their aims and audience. One of the projects made heavy use of simulation models, which made it particularly difficult to evaluate.

Andrew Foster also looked at a range of environmental projects, and had a somewhat more negative overall assessment. He wrote:

“The standards of inference on this work seem low and the extent of theoretical insight provided quite limited. This may in part reflect the special challenges of work on environmental issues as well as the relative scarcity of work outside the Bank by economists in this area. Nonetheless, many of the issues being addressed are of first order importance in terms of individual well-being. Moreover, these issues also seem to be of high salience in terms of public policy given the importance of external effects in this arena that are difficult to internalize through market mechanisms.”

Flagship reviews: pensions and insurance, Doing Business, and transition

We have already discussed at some length Peter Diamond’s review of Bank research and dissemination on pensions and insurance, and we need not repeat the material here. Much the same (in the other direction) can be said of Antoinette Schoar’s extremely positive review of the *Doing Business* work. Jan Svejnar provided an evaluation of five flagship publications on the transition, two of which are on poverty and inequality, and two on corruption. These publications were very widely disseminated and enormously influential in the transition countries. Indeed, given the previous isolation of many of the policy makers in the transition countries, and the fact that they were looking to the west for guidance, Svejnar says that these publications were treated as gospel. In the two sets of two reports, his assessment was that they improved over time, although from a relatively low base. And although he gives all five reports a generally positive assessment, one criticism is that they were relatively insulated from the large amount of research that was and is going on in the substantial number of research institutes in the countries themselves.

Given his long involvement and high-level connections with policymakers within the Czech Republic, Svejnar is in an unusually strong position to assess the effectiveness of the dissemination of this work. In particular, there is a question of how such long reports (one is 524 pages) can be so effective in reaching policymakers who are famously unable to read memos that are as long as a single page. According to Svejnar, “what the Bank usually did was to arrive in the capitals of the transition economies to present these reports, invite advisors to ministers and other influential individuals to serve as discussants, and also invite press and academics. This induced the discussants (and others) to read the reports and the reports were also aired in the form of front page articles in the press. I experienced this a couple of times when I lived in Prague and there usually was a discussion of the main ideas of these reports among the “intelligentsia”.”

Chapter 5. What we learned from the interviews

Introduction and Overview

Our team contacted a large number of people in the Bank both on the research side and the operations side, and asked them whether they felt that World Bank research has achieved its two main objectives: (1) the generation of new knowledge on development? (2) a broadening of the understanding of development policy. Many of them replied at length, often going beyond these specific questions. We also conducted a number of more open-ended interviews with some of the senior staff of the Bank, including most of the regional Chief Economists and a few current Vice Presidents, as well as a number of leaders of Bank research from the recent past, who have since left these positions. It is important to note that we could not interview everyone who might have been relevant, nor could we ensure that the opinions that we heard were representative, or even well informed. Indeed, it is quite possible that some of those who were keen to talk to us were people who wanted to take the opportunity to express their unhappiness about research in the Bank.

Finally, we talked to a number of senior academics in developing countries, as well as a number of leading policy-makers and policy advocates outside the Bank, with the aim of finding out whether they thought that Bank research was useful. We had no good way of doing this systematically, and in the end, heard from only a handful of people.

It is important acknowledge that the responses we have or even the list of people we contacted was in no way representative of some easily identified target group. In many cases we relied on our personal contacts to get the interview and in others, our rationale

for contacting the person was based on our perception of the person's prominence in the Bank or outside.

We told our respondents that their names would not be used in our report and that we would do what we could to protect their identity. Many of them were actually prepared to let us use their names, but in the end, we decided against doing so. In the case of interviews we present material as if it were a direct quote, but it must be remembered that the quotes are from our transcriptions of what was said and should not be treated as literal.

In spite of all these limitations, we feel that the interviews and messages that we received yielded a great deal of information, and offered us a unique perspective on how Bank research is perceived and the hopes and expectations faced by Bank researchers. We therefore decided to include a summary of the interview material in the report and it is included in the body of this chapter. Here we provide a short summary of the main conclusions.

The first message, which came from all sides, is that the Bank needs to do research in house. It cannot be completely contracted out to academics without undermining much of what is uniquely important about it. Most, including those from operations, also supported the idea that there should be some scope for doing "blue skies" research within the Bank, partly on the grounds that it is an intrinsic part of doing research.

There was more disagreement about whether Bank researchers were striking the right balance between research quality and relevance. The view among most people from research seemed to be that maintaining the right balance was difficult, but things were not very far off the mark. The view from operations tended to be more negative. Our

respondents felt that while there many individual pieces of useful research, there was also a clear gap between what they wanted and what the researchers were delivering. A lot of the research did not help to answer questions that operations wanted answered, even when it was in the right area, in part because relevant work might be less gratifying from the research point of view. The lack of relevance was particularly a problem in countries where it was difficult to work, such as African and Central-Asian countries. It was very hard to get researchers to focus their research on issues that loom large in these countries.

Both past and present leaders of Bank research emphasized that this conflict with operations was in part inevitable, since one important role of the research department is to keep operations “honest” by making sure that what operations is pushing is in line with the best of current research. Indeed several expressed the fear that researchers are often too tame in the way they carry out this particular role because of a fear of repercussions within the Bank, and emphasized the need for more independence. Several outside users of Bank research from the policy world also felt that bank research often seemed to fall in line with what the operations people were pushing, and for that reason, policy-oriented products of the Bank research department such as the WDRs and the PRRs often tend to be one-sided and/or bland.

However many people, both inside and outside research also expressed the view that the internal incentive mechanisms of the Bank were ineffective in directing research. The heads of research, including the Chief Economist and the Research Director, often had relatively little influence on what research got done, in part because the committee in charge of the allocation of research funds (the Research Committee) was reluctant to take stand on issues of relevance, and in part because Bank staff, in effect, have “tenure”. In

addition, people from the research department also pointed out that a lot of research in the Bank happens outside the research department and therefore is less clearly under the control of the Chief Economist.

The view from inside: how World Bank research perceived by Bank researchers?

Current bank researchers, whether or not they are formally in the Bank research group, are, perhaps predictably, emphatically positive about the importance of the Bank being involved in research. Many of them are very proud of the best research done at the Bank and the impact it has had on how the rest of the world thinks about development.

However they are also concerned that people should appreciate the special role played by Bank research in sustaining operational work and warn against simply valuing the research based on the publications it generates. As one of them puts it:

“The real work of DEC is often unobserved, and consists of making arguments in meetings, providing an impartial voice on internal debates, providing technical advice, etc. This is an invaluable role.”

It is therefore no surprise that with one exception, they emphasize the dangers of thinking that the research that is now done in the Bank could be contracted out. And the one person who does suggest contracting out a lot of the research admits that while there is a theoretical rationale for DEC it has rarely worked out in practice.

The general sense seems to be that things are more or less right, and several warn that there are dangers in trying to fix what is not broken. At least one of them goes even further, writing:

“Finally, although the email solicitation asks for ideas for improvement, I have to demur. The Research Group is open to and has made significant changes in the past in order to increase relevance and quality. The Group's exposure in Bank budget battles forces it to constantly defend its worth. Bank research is evaluated every three years by the Research Committee. So there is an institutional reason why we might be more or less close to the production possibility frontier.”

However most recognize that there are problems. Several mention lack of resources to collect data and hire new PhDs so as to guarantee a regular injection of new ideas into Bank. Some complain about the demands of doing cross-support work in operations, while several mention the connection with operations as one of the main advantages of working as a researcher in the Bank.

A number of leading researchers complain about the emphasis on putting together high visibility reports such as the World Development Reports (WDRs), the Global Economic Prospects (GEP) and the Global Monitoring Reports (GMR) and point out that they are extremely intensive in researcher time (the WDR itself absorbs 8 full-time researchers for a year, which is about 10 percent of the research group). One of the researchers goes on to say

“I also think that Policy Research Reports are a much more flexible and cost-effective vehicle for disseminating research to policymakers. They cost a fraction of what WDRs, GEPs, and GMRs cost. The budgets for them finance new research that gets incorporated into them, rather than just being literature surveys. Importantly, they are produced "on-demand" rather than on a fixed timetable. And my guess is that impact per dollar spent for some of the better ones, such as the ones on the East Asian Miracle, Aid Effectiveness, Fighting Aids, and others, could be much higher.”

However another leading researcher calls the WDRs the most important thing that the Bank does and argued that they had a large effect on thinking, inside the Bank as much as outside of it..

Another issue that comes up is that of the multiple audiences that Bank researchers may satisfy. On the one hand, Bank researchers have to take care to frame their results in a way that will pass serious professional scrutiny, often in highly critical and competitive refereed journals. The editors and referees of these journals are typically tough on papers that overstate their claims. At the same time, Bank researchers have to satisfy their operations counterparts who want to see research that has clear policy implications, ideally in line with policies they are already espousing.. As one Bank researcher puts it:

“Operations may often protest that research stands in the way of doing work when findings which question current best practices are presented. This may lead Operations staff to feel that research is not useful. Research which is supporting what is happening is more easily integrated into operations, but it does not mean that research is actually having an impact in this case.”

There is however disagreement about what to do about this. Some suggest scrapping the two publications requirement in order to allow people to do useful work, such as working on a country report. Others resist, arguing that:

“The requirement that staff publish 2 papers each year is necessary to ensure that the quality of WB research remains high. Publication is a way of generating credibility, which leads to impact. If you look at institutions which don't have this requirement, their research does not have as much credibility.”

There is also some sense that the current mode of integration of research with operations may not be optimal. One researcher complains that there are no opportunities to integrate

randomized experiments or even base-line surveys into Bank projects. The DEC data group also complained about the lack of integration between research and data collection.

Finally there is some concern among DEC researchers about research in the Networks. The perceived problem comes from the fact that the research in the networks is not necessarily subject to the same level of scrutiny as the research coming out of DEC and may be of a lower quality, and more subject to ideological manipulation. One leading researcher in DEC says that the networks are “out of control” and another is even more blunt:

“Research outside DEC - with a few notable exceptions - is even more practical. It is essentially a form of rhetoric. It is often not about doing research to discover new knowledge but to justify some previously determined policy. It is not unusual to be told that "we should do an evaluation to prove that X program works," for instance. Or "we have to run some regressions to show that Y agenda matters for growth otherwise we will not have Bank buy-in." Peer reviewing is often fixed by appointing cronies as reviewers who are not in a position to make critical comments.”

On a more positive note, one of the researchers in a network described a process he is putting in place to help identify research opportunities and in particular opportunities for randomized evaluations in the Bank’s project work. These opportunities will then be offered to DEC researchers, though he expressed the concern that this initiative might collapse if he were to leave the Bank. Another researcher outside DEC points out that

“Two of the most-often used pieces of World Bank analysis are not generated in DEC: Kaufmann's corruption indicators and the Doing Business indicators.”

Looking back: The views of some past leaders of the research department

We talked to a number of people who, in one way or another, were in a leadership role in the research department in the past and have now moved on. We felt that this puts them in a unique position to comment on the constraints facing the research department in the Bank. As it turned out we were very fortunate to get a number of very engaging responses.

All of our respondents recognize the value of doing research in the Bank. A number of them emphasize the importance of independent research that is not necessarily aimed at making operations happy. As one of them pointed out

“Another central issue in evaluating the relevance of research in the Bank to operations is whether to accept as given the way that operations function. It is possible that operations could dislike research, and yet the researchers be right, and operations wrong. There is tremendous pressure in the bank to lend, and if researchers look at a proposed project, and say that it is no good, or there are problems, they will not always be listened to.”

Our respondent goes on to suggest that this conflict may be structural, given that so much of operations is aimed at lending:

“Lenders do not have incentives to provide loans for economically sound projects. DEC may be dedicated to development in cases where operations are not.”,

and concludes that

“How we think about research depends on whether we take this as given, and try to work around it, or whether we think in terms of thorough-going reforms of incentives throughout the whole Bank.”

In another place this respondent emphasizes the importance of independence, “It is important that there be no topics that are off-bounds to researchers, though this can call

for delicacy of management from time to time.” The same theme is picked up by a number of others, who also express the concern that the current institutional structure within the Bank is not necessarily ideal from the point of view of protecting this independence. One of them emphasized the fact that “the Bank will *never* evaluate what it is doing if there is an atmosphere in which the full range of criticism cannot take place. So that it is full freedom for Bank researchers that ultimately keeps the Bank honest.”

And went on to complain that

“There was an enormous amount of interference by the PR people, especially after Wolfensohn became president; research was not supposed to offend NGOs, nor to provide them with material they could use to criticize the Bank.”

Another person suggested that this is a chronic problem:

“There is too little lively discussion and criticism of Bank research. It is important to please the hierarchy and people think that criticism might undermine the role of DEC.”

Someone else who also wants Bank research to be more iconoclastic and less supply driven, goes on to suggest that this tension between what operations wants and what research shows may be partly resolved by drawing a clearer, brighter line, between basic research to advance development knowledge and policy, as opposed to research aimed at educating ministers and country specialists. In the case of the first kind of research, he proposes, the results will not be advertised in any way as institutional views of the Bank

To fund this first kind of research he feels that the Bank should perhaps move closer to the NSF model, competitively funding development economics research by both insiders and outsiders.

On the other hand, several others we talked were explicitly opposed to going to the NSF model. The concern was that it was important for the Bank to build up a coherent knowledge base on a specific set of subjects, whereas NSF-style funding leads to a disparate set of potentially isolated research products. Indeed one of them complains that the Bank Research Committee which allocates research funds within the Bank, was *too* much like the NSF:

“The Research Committee refused to judge pertinence, but instead emphasized quality, or its predictions of quality, which in many cases they were not competent to pronounce upon. So it was very hard for the management of DEC to actually influence what was done. At the same time, individual researchers have nothing to prevent them from working on exactly what they want to work on, and indeed there are incentives that encourage them to do so”

Instead, another of our interviewees felt,

“there ought to be a move towards more programmatic funding. Programs of research, perhaps based on existing DEC groups, should be funded for, say, five years at a time, and those groups should build links with the regions, the networks, and outside researchers. That they bring in top outside researchers would be a condition for their funding. A programmatic framework would offer incentives to good researchers in the networks and the regions to be involved in DEC research.”

In other words, more directed research led from inside the Bank, but with better coordination with leading academic researchers.

There was also some discussion of whether the Chief Economist of the Bank is in a sufficiently powerful position to shape Bank research, especially from the point of view of quality control. One important concern was with research outside DECRG. One of our interviewees took the view that

“it would be a good idea to have a Chief Economist who really was a chief economist of the Bank, and not just head of DEC. So that all the economists in the Bank, in DEC, the networks, and the regions, would report in part to the CE. This would help with quality control, though it will never be perfect.”

But another demurred:

“having a Bank Chief Economist to which all economists would report (in part) would not work. These people would be sitting somewhere else, and their primary reporting requirement would be to someone else”

and for that reason our respondent doubts that the Chief Economist could exercise effective quality control.

The same respondent thinks that the only way to deal with the dispersion and quality issue is “(a) for the regions, for the CE to work closely with the regional chief economists”, and “(b) for the networks, to promote the reintegration of PREM (and the WBI) back into DEC, where they once were. For the other networks, the CE needs relatively senior people in DEC to monitor what they are doing so that it is possible to intervene at a relatively early stage.”

In any case, even the person who favors expanding the Chief Economist’s reach doubts that the Chief Economist alone can do all the necessary quality control and suggests that there may be a role for a panel of outside researchers in the review process.

Consistent with the view that the Chief Economist has too little time to do everything he or she needs to do there were questions about the value of the WDRs. One person suggested that the WDRs (and PRRs) were a prime example of research where the conclusions are “either predetermined or negotiated in advance. WDRs and PRRs are in this category.” And concludes: “This stuff is largely worthless, it does not even have an effect within the Bank, and there should be much less of it.” Another person also

expressed a similar view, albeit in a somewhat different tone advising that “the panel should think hard about whether the WDR should continue. Has it remained original enough and vital enough?” The sense was that it had drifted from a research vehicle to more of an official document. Yet another commented that “the WDRs are most effective when leadership is handed over to a smart and able person charged with giving the report their (team’s) own unique voice (“make it sound like yourself”). WDRs are least effective when the person in charge is told to draw together different viewpoints from around the Bank and exposit them.” Finally, at least one person argued for reducing the frequency of the WDRs.

The last issue that came up several times is the appropriate role of the DEC data group. One person argued:

“..there is a big *institutional* problem with DECDG and DECPG (prospects group). Once a department like the data or prospects group is set up, it has to develop its own products, which may or may not be useful. It tends to give the lowest priority to collaboration and support for DEC (and research more generally), which ought to be one of its main reasons for existence. The LSMS was as successful as it was because it was done within the research group, not by the data group.”

Another person suggested that the Data Group “needs more statistical capacity than it currently has. It is *managing* the ICP project quite effectively, but is contributing relatively little to it intellectually.”

The view from operations

Operations represent the first level of clients of Bank research. Without interest from them, it would be much harder for researchers in the Bank to have much of an impact. On the other hand, they are also the people who depend on research to deliver support for

the policies that they are recommending. If research refuses to deliver, say because it is not what the evidence points to, operations necessarily may not feel very positive about research.

The responses from the operations side to the surveys/interviews reveal wide-spread support for the idea that the Bank should do research. Several people went so far as to spell out the reasons why doing research at the Bank makes sense. One of them, from operations, who is rather critical of the performance of the DEC Research Group, starts by saying that “Research is undervalued but central to the achievement of our goals.” and then goes on to suggest that “the rationale for why the Bank needs to do research and why we need a central research capacity of the DEC type” are in part “the standard arguments in the literature for public research institutes” and in part “the need to maintain the human capital of the staff and the currency of analytical approaches found in operations work”. Among arguments for Public Research Institutes he/she mentions the “provision of at least three public goods”, including “Authoritative and Independent Establishment of Standards (for measurement, etc.)”, “Diffusion of Knowledge” and “Coordination Tasks (for combining inside and outside expertise, etc.)”.

There are certainly some who also feel that the researchers in the Bank are doing a very good job. One person from the senior management of the Bank expressed satisfaction with DEC’s work and gives this example of the impact of the research:

“One example is the research on service delivery, where research started documenting cases of failure of service delivery, e.g. absenteeism of doctors, or schools being built without teachers. This led to shift in operations. Now, nowhere in the region does the WB do stand alone education projects –all education programs are programmatic, provide budget for programs and then provide incentives for service providers; indeed loans are conditional on tackling

absenteeism. That has been a nice link back and forth between operations and research.”

Many others praise the poverty work done by the Bank. More than one person also praised the Bank’s role in helping countries understand how to measure poverty.¹⁰

However there is also a large group that feels that Bank research is nowhere near what it ought to be doing. Most, though not all, of the respondents are willing to grant that the research group is generating new knowledge on development (our first question in the survey), but are less convinced that this leads to a broadening of our understanding of development policy (our second question).

More specifically, they complain about the lack of relevance to operations work. One senior operations person reported that “Bank research is virtually invisible in ECA (the Europe and Central Asia Region).” When probed he/she explained that one major problem was the lack of regionally specialized research. Many others who work in the regions echoed the same thought.

One reason for the lack of regionally specialized research was explained by a senior operations person in the Latin America (LAC) region, who is otherwise relatively positive about DEC research: “The research at DEC is focused on priorities that are determined for the world as a whole and these do not necessarily match the priorities of the LAC region.” As a result, this respondent goes on to explain, the Latin American Region does its own research: The Chief Economist’s office acts as an intermediary, gathering research questions from the operations side and finding researchers to do the

¹⁰ Though there was also at least one person who commented that the reason why countries want poverty measured is to keep the Bank happy, so this could be seen as a case of supply creating its own demand.

work. The fact that the LAC regional office works particularly well in this respect comes up in a number of other comments as well.

The rest of the regions do not however feel that they are in a position to take over the task of supplying the necessary research. A senior economist in the East Asia region explained why:

“it has proven to be almost impossible for regional staff to do research. The reason is straightforward – DEC staff get paid to do research, while operations staff do not have their salaries paid for and have to sell services to the country directors. Country directors get the budget to do country programs and this excludes research.”

The question therefore is why DEC is not delivering the kind of research that the regions want. While LAC may not be a priority for DEC, the Africa region, with its enormous economic problems, would seem to be a natural candidate for a priority area. Yet the reactions of some of the people from the Africa region seem particularly negative, in spite of the fact that Africa is the region that is the heaviest user of cross-support from DEC.

The problem with Bank research on Africa, we heard from a number of people associated with the Africa region, is that it is difficult to get good researchers to work on Africa, in part because the data are not as good, in part because they feel less attracted to the countries. A senior Bank official acknowledged:

“The Africa region has always had a problem to attract the best managers, researchers, only got people from edict. That doesn’t produce the optimal kind of incentive structure. Easy to get people to work on Asia, Latin America.”

What makes matters worse, one respondent points out, is the fact that there is very little turnover in the research department and hence very little new hiring. As a result, you cannot typically hire a new person to do what you want done.

Apart from the lack of country specific studies we also heard complaints about the methodological biases of the DEC researchers. One of our respondents complained about

“Cool tools over relevance. DEC research often appears to duplicate academic research in its focus on applying sophisticated empirical methodology at the expense of addressing more policy relevant questions. DEC output (in my field), taken together seems to consistently make this trade-off.”

More specifically, many explicitly complain about the excessive use of cross-country regressions, which they claim do not inform policies vis à vis individual countries. One of our respondents goes so far as to say

“I’m not too sure that the outpouring of cross-country growth regressions by economists inside and outside the Bank proved anything but you can have the same arguments we used to have in prose using baroque models and estimation techniques.”

On the other hand, a number of senior people from operations explicitly come out in favor of letting people in DEC do a certain amount of “blue skies” work, on the grounds that you cannot have a research department without it, though one person explicitly argues that:

“Blue sky research is probably not the comparative advantage of the Bank – should engage with academic community for that.”

A number of respondents also feel that Bank researchers do not do enough of work where they have a comparative advantage, such as in project-based work and putting together data sets. One of our respondents remarks that

“There should be far greater institutional incentives for researchers to use individual Bank projects as a research vehicle --indeed, for Bank projects to be designed more frequently as being suited to research. The institution hands out \$20-odd billion in money each year to a group of countries with GDPs worth in the region of \$7 trillion, it has to be through very high

leverage that such resources make a difference--and one way to ensure leverage is to build up the knowledge transfer that occurs with projects.”

Another comments that

“The research on poverty seems a little esoteric, especially in view of the fact that we haven’t made the investments needed to conduct and process consecutive, compatible household consumption surveys in more than a few African countries. Thus, several years into the HIPC process we struggle to document any progress on poverty beyond HD indicators.”

Finally, there are a number of complaints about the coherence of the Bank’s research agenda. The basic criticism is that the research consists of a number of disparate pieces that do not always build on each other and aim to create a complete picture. A particularly forthright statement of this view comes from a long-term Bank operative who says

“Young and eager professionals entering the WB when given a task of preparing a project or ESW are lost and lose much time in reinventing the wheel. They miss the experiences and research built by Bank’s staff over decades, sometimes at a great opportunity cost for the client countries. Little institutional memory exists. If the WB were a consultancy firm, it would probably have been bankrupted long ago.”

The view from outside: what did we hear from policy people and senior academics in borrowing countries?

As we say above, we have responses from relatively few policy people. Among them the most interesting are perhaps the developing country policymakers, since they are meant to be the ultimate clients for a lot of Bank research. All of the ones we talked to were positive about their experience with Bank research and said they do make use of it; such views were echoed by the researchers in the same countries. Many of them emphasized the value they get from the fact that the

Bank distills current research and puts it out in the form of WDRs or flagships, which helps them stay abreast current research and also provides something that they can pass onto their staff. One person did however say that the WDRs and flagships are not very useful for middle-income countries and tend to be too general. A number of them also read Bank working papers and praised their quality. They also told us that they often directly consult experts from the World Bank for know-how/technical advice. Two of them mentioned help from the Bank in doing randomized evaluations.

However one of them was also critical of the way the Bank delivers policy advice. In particular he says that the Bank does too little in the way of cautioning people on all the many reasons why policy conclusions may be subject to qualifications or depend on the specific circumstances of countries.

A number of people involved in policy advocacy also echoed the idea that the Bank often pushes its current policy recommendations too hard. In particular, the Bank's advocacy of Defined Benefits Plans and private investments as a part of a social security program was noted by several people as being largely driven by ideology, without concern for the particular circumstances of the country.

This group of people also questioned the Bank's strategy of using thick volumes (WDRs, PRRs) to disseminate its messages, since very few people in positions of power have the time to read them. On the other hand several of them mentioned that they themselves use the WDRs. They also mentioned the World Bank as an important source of data but perhaps one that is no longer as useful as it used to be.

Chapter 6. World Bank research: exploring institutional options

The main focus of this panel is on evaluating the Bank's research *output* and on how research feeds into policy. Although issues of *process* have already emerged, we have not focused on these until now. In this chapter, we draw on suggestions by evaluators and interviewees to develop a number of suggestions on how, institutionally, the Bank might set about further raising the quality and impact of its research. In presenting our recommendations at the end of this chapter, the panel acknowledges that managing research and researchers in the Bank environment, with its diverse and ever-changing demands, is an extremely complex one. "One size fits all" formulas for producing important development policy research are just as ill-advised as for development policy itself. Nevertheless, we feel it useful to offer constructive ways to deal with some of the Bank's challenges.

Problem areas

Having discussed both the evaluators' and our own assessments of research in Chapters 3 and 4, as well as the views of a broad panel of interviewees inside and outside the Bank in Chapter 5, we are now ready to present some of our conclusions about the most significant issues facing the Bank research program.

Budget squeeze

For a variety of reasons, including a cyclical decline in the Bank's revenue from middle-income loans, the entire Bank administrative budget has been squeezed over the past few years. This funding issue has presented challenges throughout the Bank, but particularly in research, which is heavily dependent on maintaining a constant flow of new ideas. As one example of the problems that have arisen, our evaluators observed that due to a budget squeeze over the past several years, the Bank's core research group DEC has only been able to engage in very minimal hiring of new PhDs, even at a time where the quality of new PhD students in development seems to be cresting.

At a deeper level, life can be very difficult for advocates of basic and applied research at the Bank, despite the Bank's rhetoric in recent years about its new role as a "knowledge bank." The fact is, that despite the very high payoffs to research, the long gestation periods make it enormously difficult to maintain a research program on a scale commensurate with the Bank's overall role in the world of development. Research is seldom part of an income-producing lending program, and it often fails to deliver the simple syllogisms that management wants to espouse in its advocacy of what it believes to be best development practice. The simple and compelling fact is that despite its centrality to the Bank's mission, research only accounts for 2.5 per cent of the Bank's administrative budget. By contrast, and as an example, the fraction devoted to supporting the Bank's executive board is more than twice as high. The future of the Bank's research function depends on developing a more stable financing mechanism; we will discuss this issue further below.

The Bank should be able to produce a lower proportion of research that is neither policy relevant nor academically distinguished

In Chapter 3, we noted that there is a great deal of excellent research coming out of the Bank, albeit not always the research that receives the most prominent recognition. At the same time, the overall impression of the panel and the evaluators is that there is altogether too much undistinguished and misconceived research coming out of the Bank. The panel certainly appreciates that a good research environment allows risk taking, and accepts failures as inevitable consequence of risk. Moreover, a few great ideas can justify even a bevy of failures. It also understands the fundamental tension that comes from requiring academic publication in an organization that is fundamentally concerned with policy. While the publication requirement is necessary to maintain quality, it inevitably leads to some research that that “academic” without being either very original or relevant to policy. However there seem to be many instances of weak research that could have predicted ex ante. For example, a lot of the work on civil wars, as evaluator Daron Acemoglu points out, builds on a combination of theory that is 25 years behind the current state of knowledge and empirical work that is extremely deficient. Projects that propose to use flawed techniques, or inappropriate consultants, should either not be approved or should be terminated quickly.

What is clearly needed is a stronger mechanism for giving critical feedback and, ultimately, for cutting off funding to projects that are unacceptably weak. Strong mechanisms can keep the bottom tail of research from becoming too big and too long, as we find it to have been over the review period.

The panel recognizes that the “weak tail” problem has many causes. As noted in Chapter 3, there are some fields (such as urban economics or infrastructure) where academic research is especially mixed, and where the Bank is forced to take very big risks in pursuit of its objectives. But there are also some glaring institutional problems. For example, whereas the Bank needs to encourage research outside the research department, it is also needs to develop procedures to ensure that all research is subject to strong and critical feedback mechanisms.

The fundamental tension between the Bank’s role as an advocate of good policies and a producer of new policy ideas

One of the biggest tensions faced by Bank researchers is that, on the one hand, they are asked to produce evidence showing that Bank recommended policies – presumably best practice – actually work. In this sense, research is a lynchpin of Bank credibility. On the other hand, researchers are expected to come up with bold new ideas that, inevitably challenge the status quo, and therefore question existing Bank practice. Overall, Bank research has done a credible job balancing these two roles. But in recent years, the Bank’s policy arm has faced enormous political pressure from other international organizations that make no pretense to balance in their anti-poverty analyses. These organizations offer theories of how poverty can be reduced without any serious consideration of the evidence that (often sharply) contradicts the positions they advocate. Instead they take support by celebrities and rock stars as a substitute for cold analysis. Admirably, the Bank has attempted to resist this populist temptation, realizing that in the

long run its credibility will be shattered by one-sided advocacy devoid of suitable objectivity and balance.

Unfortunately, however, the enormous pressure of populist rhetoric has taken a toll on the Bank's ability to balance advocacy and objectivity in presenting research results. And this is the area where our review uncovered the most widespread and troubling issues. Enormous problems can occur when not-very-robust research results are sold as irrefutable truths to the countries in the form of policy advice, technical assistance or as part of the conditionality of the lending programs (in Chapter 3, we gave examples in the area of pension reform, financial sector liberalization, aid effectiveness, poverty mapping, and the effect of globalization on poverty). Even when the underlying research is valid, the Bank's desire to get out a message through external communications can give the impression of crisp black and white results, with too many important nuances lost.

The balance between rigor and relevance

One of the great challenges in managing a high-level policy research program such as the Bank's is how to balance the need to give researchers leeway for creativity while at the same time creating clear incentives for them to deliver ideas on the topics the Bank cares most about. The challenge for the Bank is how to nurture a top-level research environment – which necessarily implies housing researchers who keep reasonably close to the frontier on new methodologies – while at the same time ensuring that an adequate percentage of research time is allocated to policy relevant projects.

Recognizing these difficulties and tensions, the panel and evaluators still find that there seems to be an inadequate connection between the demand and the supply of development knowledge. On the one hand, a great deal of Bank research may pass the high standards of academia but be irrelevant to the World Bank's clients and mission. On the other hand, some of the most relevant questions for the countries and the global economy are not addressed by World Bank research. The panel acknowledges that this is a difficult management issue; the best research ideas tend to come from the bottom up rather than the top down. Often, ideas that at first appear very abstract turn out to have extremely important practical implications. The returns on Bank research must be evaluated as the output from a necessarily speculative portfolio. The fact that high-risk/high-expected-return research projects do not always succeed should not be viewed as a deep systematic problem. Nevertheless, the panel believes that finding ways to improve communication between clients and researchers would energize research at the Bank rather than eviscerate it.

Balance between responsiveness and independence

A different dimension of balance comes from the need to be responsive to the needs of operations while maintaining the appropriate intellectual standards. Operational staff want answers to the specific policy questions that they are currently tackling. It does not necessarily want to be told that there are no reliable answers available right now and that it might take years to make progress on the question. On the other hand, researchers may be unwilling to take a stand on the issue, given that the research is really not there. The extreme version of this is when operational staff want a *particular* answer which does not

happen to be consistent with what the best research is showing. For the World Bank's advice to be credible, it is important that in such cases researchers be in a position to make their reservations heard and have an influence on the outcome. We have the impression that researchers in the Bank are not always sufficiently insulated from pressures from operations to have this kind of independent influence.

We also heard about cases—possibly not widespread—where independence may also be a problem vis à vis the leaders of the research department, who might also want research that delivers a particular outcome. The value of research, of course, comes from the fact that you do not always get the answer you want. This tension is therefore endemic and requires a great deal of forbearance on the part of the research leaders and managers.

Data collection and maintenance

There is widespread agreement that thoughtful data collection, guided by researchers who understand what types of information is most needed to address fundamental policy questions, is one the Bank's greatest contributions. The Bank has taken giant strides in improving its work in this area over the past ten years. Nevertheless, despite many promising steps, data collection at the Bank, as well as access to data, remains extremely haphazard. Not all the databases collected by the Bank are archived and maintained, nor are most disseminated by the data group. Access to data is further sometimes undermined by technical constraints (no user-friendly software available) as well as bureaucratic constraints (national governments often allow Bank researchers to use national data on the condition that they sharply restrict access to others). A cost-effective

mechanism needs to be found to make the most important data sets well maintained and easily available. Even when data and documentation are available, the Bank website is often a very model of unfriendliness; important links do not work, and search facilities do not find documents that later turn out to be present. Furthermore, and until very recently, the Bank has not done as much as it might have to push for the greater international harmonization of the survey data that are to be used for measuring improvements in living standards.

Statistical and econometric expertise

The panel has a general concern about the Bank's collection and management of empirical evidence. Over the last twenty years, there has been an enormous change within the Bank in the way data are handled, and in the use of data in empirical research and policy advice. Many Bank projects collect new data, and the vast majority of Bank research uses econometric methods that were essentially unknown in the Bank as late as 1980. Yet the corresponding changes in the provision of statistical support and management have not been made. The Bank provides no central survey support organization, so that when it comes to survey design, sampling, and questionnaire design, researchers are entirely on their own, or at least in the hands of consultants. It was many years before the LSMS team understood the difference between a self-weighting and a simple random sample. But statisticians in the DEC data group are rarely consulted by Bank researchers, and even within DEC, it seems that there is little real intellectual interchange between DECRG and DECDG. Both data collection and data analysis suffer from that situation. As far as we can tell, Bank surveys are not subject to the human-

subjects protections that are now standard throughout academia, even in research that is not related to health; this situation represents an accident waiting to happen.

It is also our impression that the management has not always kept as far ahead of the methods being used by researchers as would be ideal for effective management. Although this is a Bank wide issue, DEC is the obvious home for statistical expertise and quality control in the Bank. Some of the most conspicuous problems in the evaluation concerned the overstating of conclusions—sometimes in high-profile reports—based on flimsy and/or flawed empirical evidence, not exerting appropriate oversight over poor empirical methodology, and allowing projects to run for many years without making sure that they were on sound statistical foundations. Better oversight of methods and techniques is likely to help Bank researchers come closer to their potential and will help shorten the long tail of undistinguished research, even without changes in the rules that researchers face, for example in the publication requirement.

Too Many Thick Volume Flagship Reports

The panel appreciates the role that thick volume flagship reports can play in energizing and publicizing research around important topics such as pensions, urban development and health. However, there does seem to be a sense in which the Bank produces altogether too many such reports, resulting in very uneven quality and impact. We have already noted the huge resource cost of these reports. Of far greater concern, however, is that the plethora of flagships makes it virtually impossible for management to exert sufficient quality control. The Chief Economist's office, even if it were vested with sign-off

authority on all flagships outside of DEC, lacks the time and resources to adequately vet them.

The issue of quality control is of great concern because flagships are typically vehicles where the line between the Banks' advocacy role and its role in producing new research ideas becomes particularly blurred. In Chapter 3, we noted the profound problems with the flagship report on "Assessing Aid", where policy recommendations were based on a very small body of preliminary research results that ultimately proved unsound. This is far from an isolated problem.

At the same time, the panel found that many Flagships, particularly over the more recent part of the review period, tried to please too many constituencies, and as a result, lacked sharpness and focus. As a result, they added too little to the development debate, a point also noted by some of our outside interviewees. This is even a problem with the World Development Reports. As one former Bank chief economist noted, the weakest WDRs seem to be those that attempt to synthesize a broad panorama of ideas from around the Bank, instead of focusing in clear on a particular clear and original theme.

It is also at least questionable whether the Bank should always rely so heavily on flagships as a vehicle for disseminating research ideas to policymakers. It is notable that a private consulting company such as McKinsey, which sells its technical advice rather than dispensing it freely as does the Bank, tends to rely more on much shorter presentations (sometimes consisting mainly of bullet points and diagrams) in conveying its messages. While this is clearly not a model for the WDRs, it might be worth considering in other contexts.

More Support for Institution Based Research in Developing Countries

The Bank provides some support for institution-based research in developing (client) countries; important examples are the African Economic Research Consortium, DEC's large and continuing support for the Global Development Network, and more limited support for some organizations such as CERGE in Prague. But there has been relatively little involvement of country researchers in the mainstream research of the Bank, for example in the projects that we reviewed. This is a problem both from the perspective of having new research ideas filter into developing countries, and from the perspective of losing an essential source of new ideas and data sets. By providing more resources for joint and for institution-based research in client countries, the Bank would gain several benefits: (1) research would be naturally focused on issues that are important to the countries; (2) the Bank would contribute to the strengthening of research institutions in the developing world; (3) local research institutions are often one of the best conduits for policy ideas, as their ranks often have close ties to governments, so this could also be an effective way of strengthening advocacy for good development policies, even when the main research effort comes from Washington.

One of the evaluators (Nancy Birdsall) mentioned that the Bank could take the research networks sponsored by the Inter-American Development Bank (IADB) as a model. There are a number of top researchers from developing countries currently working in upper income countries (particularly the United States) who might be enticed to spend more time in their home countries if more funds were available to support their research.

Addressing the problem areas

We now turn to a brief discussion of how some of the issues raised in our evaluation might be addressed. We give more detail in areas that relate closely to the material covered in Chapter 3, whereas in other areas we simply sketch the general issue.

An overarching recommendation: Learning what works and telling the world

Perhaps the most important role of Bank research is to learn about what works, and to widely disseminate the results. Research is the key to quality control, and research will only be as strong as the Bank's commitment to quality control in *all* of its activities. Researchers must be involved in operational work from the beginning until (after) its end, and every project and policy should contain the tools for learning from it. We believe that the Bank should make still greater use of randomized controlled trials than it currently does, and we welcome the initiatives that are currently under way, including the Development Impact Evaluation (DIME) initiative. But much of the Bank's current portfolio cannot be evaluated in this way, and it is important that other methods of learning be strengthened. Theoretical analysis is important, to guide experiments when they are possible, and to interpret and understand outcomes when they are not. Data collection and empirical analysis are also vital tools for learning. Most of our recommendations speak to how Bank research might deal with these areas.

Financing Research and protecting its independence and objectivity

Many of the other problems we identify ultimately trace to the need to continually lobby for research, and to protect basic research on development issues, especially when the

payoffs are not immediate. The panel is particularly concerned with finding a way to fund Bank research that protects its independence, and guarantees that Bank research does not degenerate into pure advocacy of the type that has become all too prevalent in the global poverty debate. To address these two problems, our favored solution would be to create an endowment to fund the Bank's development policy research, which could be done using a small fraction of the Bank's massive cumulated retained earnings. (There is a case to be made for funding the Bank much more broadly in this way, but the case of research the need is especially compelling.) Of course, institutional mechanisms would need to be created to ensure proper oversight and control, as well as to deal with a host of technical issues such as how to share overhead, but these problems are routinely solved by universities and other organizations and could also be solved within the Bank. However, in designing such a scheme, it is important to maintain a system where researchers are pushed, not only to justify current management and board initiatives, but also to provide critical feedback.

The panel recognizes that Bank research attracts a certain amount of funding from outside donors and this could continue, provided these funds do not excessively distort the Banks' mission.

Regardless of how the budget squeeze problem is resolved, the Bank also faces an important competitive challenge in attracting and maintaining top researchers. Salaries and status for research managers and researchers at the World Bank have suffered due to internal reforms at the Bank while, at the same time, salaries at competing academic institutions have exploded. Management should review the pay-scale as well as the terms of reference of staff in the research department to ensure than incentives (pecuniary and

nonpecuniary) are well-aligned with the institution's objective of doing top quality research in development issues.

Control mechanisms for more consistent pruning of weak research

Certainly the Bank could benefit from resuming periodic overviews such as the present one. The panel notes that there has not been a serious outside review of research since 1998; research *proposals* considered by the research committee are reviewed by outside academics (though none of the panel can recall ever being consulted), but there is no external review of research output. We believe that if a review of the current kind had been conducted earlier, it might have been possible to cut off or repair at an early stage some of the weaker research projects we identify in Chapter 3. But more systematic improvements in Bank research require having a larger fraction of output be subject to a better review system. An obvious idea would be to bring more research back under the general guidance of the Chief Economist, whose office still provides the best quality control at the Bank.

The current quality control mechanisms simply have too many holes. In particular, a large proportion of Bank research is not subject to the quality control of the Research Committee or any other form of quality control that involves the Chief Economist. This is basically true for all non-DEC research. Peer reviewing mechanisms by external renowned academics or policy researchers (from the "North" and the "South") not only at the proposal but also at the output stages might substantially reduce the "tail" of weaker research. (We note that the extent to which the Bank uses outside consultants makes it difficult always to avoid potential conflicts of interest; however, we would tend to favor

competence over lack of conflict of interest when there is a clash.) Such reviews would also inject valuable feedback and discipline even into the best projects. The panel recommends that a peer reviewing process should be applied to all research whether carried out by DEC or non-DEC and whether its purpose is to produce research working papers or flagships and reports. We note that the reviewing process must necessarily be somewhat backward looking, as much as a mechanism for ensuring the continued flow of resources to parts of the Bank that have been successful as to judging the merits of prospective research. Another idea that merits further consideration is to have all policy research reports reviewed by two outside reviewers and have the reviews published at the end of the policy research reports under the reviewer's name. Compensation should be commensurate with the work required to produce a review that measures up to high standards.

Our evaluators noted that some of the Banks' best research, as well as some of its worst research, involved outside researchers. It appears that in many cases, the Bank could have anticipated what it would get by looking more closely at the research records of the researchers it hired, particularly in the case of more senior ones. Of course, at the same time, we have argued that it is important to devote research funds to institution building in developing countries. Again we would strongly favor having a system with better checks and balances on hiring outside researchers, and under which hires of consultants were subject to some minimal audit standards – again, ideally under the general guidance of the Chief Economist.

Improving Flagship Reports

The Bank should aim to produce more consistently excellent flagship reports, even if that means producing significantly fewer of them. The panel believes that it would be enormously helpful to have all flagships and reports, whether in DEC or not, be subject to the scrutiny of academic experts even if (or indeed, especially if) they are aimed at educating ministers and country specialists. Internal reviewers should also be used, with the Chief Economist's office having a particularly large influence in the process. (The Research Committee, as presently construed, seems ill-equipped to perform this role, because it is too reluctant to take on controversial issues.) Reviewers should pay special attention to the question of whether research cited to uphold policy recommendations is properly represented.

Our panel discussed the possibility of recommending that WDRs be published every other year instead of every year, in order to make them better focused and more meaningful. In the end, however, we accepted the need for the Bank to have a major annual quasi-research document, and would prefer to see the problem addressed by pruning back other flagships. Arguably, however, there is no need for a separate report giving macro forecasts, especially as the IMF and OECD produce similar reports that seem to command more international respect. (We understand that that the Global Economic Prospects flagships contain a good deal more than forecasts, but that material was not included in our assessment.)

Strengthening interactions with academics and bringing in new ideas

Managing research is a complex task and we will confine ourselves to two recommendations, the most important of which is to improve the Bank's visiting scholar program. Visiting researchers help bring in new ideas, they provide methodological assistance, and they are a conduit for students who might later come to work at the Bank.

It is very important for the Bank to have a significant program for visiting researchers. This will help continue to bring in new ideas as well as stimulate these researchers to work on topics of interest to the Bank thus multiplying up the Bank's own research efforts. Not least, a visiting scholar program should help in terms of recruiting. To this end, the panel also sees it as important to develop better and more stable mechanisms for choosing and funding visiting researchers, as well as outside research collaborators. The present system seems woefully chaotic and inadequate. The chaotic nature of the visiting scholars program is possibly due to the general budget squeeze on research. Several internal interviewees mentioned that the current budget programming strictures make it difficult to set up the visitors program in a rational way.

Remuneration for top external researchers who may wish to spend a sabbatical in DC or a country office should be determined on a competitive basis. Joint work with external academics should be encouraged but with academics who are leaders in their fields.

We also recommend that the Bank consider instituting one or more annual research conferences with a more academic orientation than the World Bank's ABCDE conference. The conferences we envision would essentially be working meetings centered loosely around a set of topics. The IMF's annual conference is an excellent role model. There is an annual call for paper submissions and many of the paper are

ultimately published in a special issue of the IMF Staff Papers. Importantly, Staff Papers may elect not to publish individual papers, and authors may publish their papers elsewhere if they choose. These meetings help bring in new ideas, and provide a useful forum for discussions. Also importantly, they are structured in a way to facilitate serious discussion and debate. By contrast, the Bank's oversized ABCDE conferences appear to have become steadily weaker in recent years, at least as a vehicle for research, and are in need of a major overhaul. They command considerable resources, are of very mixed quality, and do not seem to engender the discussion and interactions that ought to be a central purpose of such meetings.

Dealing with the Bank's overly diffuse structure for allocating and planning research

The Bank's current approach to integrating research and policy is producing inconsistent results across networks and regions. Within the Bank, some Vice Presidencies, Regional Chief Economists and Country Directors were very satisfied with the current arrangements while others were bitterly critical. However, given that the World Bank has only six regions, the fact that reactions were so disparate means that the current system is not robust.

There clearly is a problem with the "market" mechanisms in place because the "market" is not clearing: the funds available for research managed by the Research Committee are not fully utilized while some Chief Economists felt that there are very important issues that the research group(s) at the Bank is (are) not addressing (for example, issues relevant for transition economies).

Although it is clearly beyond its mandate to make a recommendation on the issue, the panel does not understand the rationale for the current departmental structure, and in particular, why the research department and at least some of the networks should be in different Vice Presidencies. Nor why it the Chief Economist of the Bank, who is widely perceived as being responsible for the quality of the Bank's economic research, should not in fact be so. Even with multiple research groups in the Bank, some outside of DEC, there is a good argument for giving the Chief Economist, as Chief Economist rather than head of DEC, the tools to exert more extensive quality control.

Some of the former managers of DEC mentioned that it would make sense to move to a more programmatic funding of the budget allocated to research in DEC. Without compromising either academic freedom or creativity, the Research Committee (or an analogous body) could identify the areas and questions of priority to the Bank and request proposals. The selection of the proposals would not be the task of the Research Committee alone but would be done in closer conjunction than is currently the case with external evaluators (see above), thus guaranteeing both policy relevance and academic soundness. The panel was struck by the comment made by former high level official in a donor country who also had a high level management position at the World Bank. According to this person, his government had succeeded in influencing the research agenda through its use of trust funds. The money was to be allocated only to the themes identified as priorities. Something similar could be done at the Bank through an appropriately augmented Research Committee (or an analogous mechanism).

Another factor that seems to be limiting the amount of country-relevant research is that in some Vice Presidencies regional staff do not have time for research. In order to be

involved in a research activity, the staff's time would have to be "purchased" for that purpose by the Country Director. Some regions (e.g. LAC) do quite a bit of research because the CDs are aware that otherwise the often highly sophisticated government officials will not listen to the Bank (and this may affect its lending program and the quality of the policy dialogue down the road).

Several interviewers from within the Bank mentioned that there is also lack of incentives (in fact, some said there are disincentives) for "return migration." If someone from Research leaves DEC to work in the operational side of the Bank he/she may have some difficulty moving back, although there have been some examples of people who have done so.

Overall, despite many of the evaluators' critiques, we view the Banks researchers, and particularly its Chief Economists, as having done an extremely responsible job of rising to this challenge. Nevertheless, over the review period, we are concerned that the independence of Bank research may have frayed at the edges. To address this problem, we see it as important both to restore the supervisory and quality control scope of the chief economists' office – which was notably clipped at the end of the 1990s with the breakup of the precursor of DEC into DEC, the World Bank Institute, and PREM. More fundamentally, however, we view it as important to give the research arm of the Bank a far greater degree of fiscal independence from the money-producing lending arms of the Bank which sometimes dominates it. This would ideally be accomplished by endowing the Bank's research activities, which could easily be done out of the Bank's \$35 billion plus in retained earnings. Before returning to this issue, however, we will try to

summarize the issues and suggestions made both by external evaluators and the broad range of policymakers, outside researchers and Bank Staff with which the panel spoke.

Making data truly a public good

Everybody recognizes the Bank's fundamental role in generating and/or disseminating data. From standard run-of-the-mill indicators to the results of randomized controlled trials, the Bank is in a unique position to make policymaking increasingly more evidence based. Bank data are vital in its own research, and are widely used by researchers, policy analysts, and governments. In recent years, the capacity to use and analyze data has improved greatly, particularly in some of Bank's client countries, so that the value of data production becomes greater every year. Bank data are (more than) twice blessed, because they support policy-making and research in the Bank, while doing the same thing in the member countries.

Despite the Bank's enormous contribution in the data sphere, it is not doing enough to ensure that there is completeness, accessibility and transparency in the information that is required for high quality analysis and sound policymaking. The panel recommends that management analyzes the role, size, budget and skill mix of the Data Group so that the Bank can be a leading institution in data generation as well as dissemination. In this process it is very important to avoid building a "data silo" and to ensure that research is integrated with data production and dissemination. The Bank may also explore collaborating with other institutions in this endeavor. A good role model is MECOVI, a joint initiative of the Inter American Development Bank, the World Bank, and the UN's economic commission for Latin American and the Caribbean (ECLAC) to improve the

quality of household surveys and the capacity of the statistical institutes in Latin America and the Caribbean. One important feature of MECOVI that the Bank should replicate is its focus in strengthening local capacity to generate high quality surveys (something which only occasionally happened with the LSMS).

The Bank's DECDG is already working with the International Household Survey Network, which was set up by the Bank and is managed by DECDG, to generate software that will help national (and private) survey organizations disseminate data and supporting documents (metadata) in standardized, anonymized, form. Such initiatives reduce the cost to countries of making data available, and should do much to make countries more willing to share their data. We strongly support this initiative, and recommend that the Bank do everything possible to help make surveys more internationally comparable. The Bank should also ensure ex ante that governments do not restrict access to survey data (issues of confidentiality can be dealt with in other ways, for example by helping countries prepare anonymized and well-documented versions of their data, and this is a prime candidate for technical assistance). Whether this is a decision to be taken by the Board as part of the commitments that governments undertake for being members of the Bank, or through conditionality in Bank's operations, or by using moral suasion, is for the Bank to determine. The panel feels strongly that if this is not accomplished, learning what policies and interventions work well and in which settings will be much slower than it need be.

The Bank's data group, which used to be mostly concerned with assembling and collating data, is now producing a good deal of original data on its own account, most importantly the purchasing power parity price indexes from the International Comparison

Project. We have already recommended that a strengthened version of DECDG be the central agency responsible for advising researchers throughout the Bank on data collection, and for storing, documenting, and disseminating the results. Researchers and other groups in the Bank should not be collecting new data without vetting and approval by qualified statisticians on a centralized basis. As this statistical function expands, management will need to develop protocols to govern its activities, for example on the timing of the release of data, and on guaranteeing that there is a single uniform and defensible source for important data, such as the PPP price indexes, the poverty numbers, and so forth. The Bank's statistical and data activities have become sufficiently important to its activities to justify the existence of something approaching a central statistical office.

Recognizing that the Bank needs feedback on where to spend its limited (but hopefully expanding) data collection program, it is important to consider approaches such as having a periodic (once every three to four years) external review panel (for example of statisticians and researchers) examine how the Bank is managing data for external research use, and helping to suggest priorities. The panel notes that DECDG already has external advisory boards for other aspects of its work, most notably the International Comparison Project.

Improved cost accounting for research

While it is true that it is hard to manage academic research, and while it is hard to attribute research budgets to research outputs, the Bank's current record keeping is below any minimal level of acceptability. Academics face the same problems of reporting on

and accounting for research, and yet are able to report regularly to their funders in a way that the Bank seems unable to do. Worse still, it was only after an enormous amount of work, and as the final draft of this report was being prepared, that it was possible for the Bank to produce a complete bibliography of publications over the last five years. A list of publications would appear to be the minimal prerequisite for any attempt to monitor and control the quality of research. The panel recommends a major overhaul of the way that Bank research is managed and reported on. This is in addition to our recommendation that the work be subject to an outside review, preferably of *all* work, every few years. The sampling that we used in this evaluation, while necessary, is far from ideal because it can easily miss projects of importance, and because it makes it hard to follow threads in Bank research.

Creating a More Formal Mechanism for Research Replication

Following up our discussion in Chapter 1 on the role of the World Bank in research, the panel would argue that the current incentive structure does not place enough value on having researchers replicate important new empirical research ideas, then applying them to other countries. The Bank should consider setting up a unit that specializes in this activity. The unit would both help assess the robustness of important results the Bank hopes to rely on, as well as expand knowledge by seeing how results obtained on, say, United States data, apply to other countries. The Bank already produces work of this type, but we feel that the creation of a unit would systematize and incentive this valuable activity. This work is clearly linked to our argument for more randomized trials and better evaluations. Although this work is unlikely to find its way into top journals, we do

not regard that as an objection. It should aim to replace some of the long tail of undistinguished Bank work to which we have repeatedly referred. The objection to this work is not only that it is of little academic interest, but that it also of little relevance to policy, something that replication studies should help address.

Summary of recommendations

General

- The research staff of the Bank needs to be seen as the main channel through which the Bank learns from its work. They need to be involved in the planning stages of policies and projects, for example by helping to set up randomized trials when possible, or in other cases by putting in place the platforms that will allow subsequent evaluation and learning. There should be a unit specifically charged with attempting to replicate promising new findings, whether from randomized trials, other evaluations, or outside research.
- Research at the World Bank should be endowed, to better insulate it from the need to constantly defend itself, and to ensure some independence from the Bank Board and management's preconceptions and prejudices about best practice policies.
- Managers of research at the Bank need to maintain checks and balances that preserve the credibility of its research. In particular, it needs to resist the temptation to make strong claims about preliminary and controversial research that appears to support policies that the Bank has historically supported.

Managing and evaluating research

- Quality control of the research program needs major improvement, with a system of regular reviews (in particular, from external peer reviewers), as well as better (i.e. some) monitoring of value for money, and reporting of outputs.
- We are puzzled by the current organizational structure and raise the question of whether the research department and at least some of the other groups and networks might be part of the same Vice Presidency, possibly supervised directly or indirectly by the Chief Economist.
- Management should review the pay-scale as well as the terms of reference of staff in the research department to ensure that incentives (pecuniary and non-pecuniary) are well-aligned with the institution's objective of doing top quality research in development issues.
- We recommend that research managers exert more careful and more central control over the quality of consultants.
- The Research Committee should consider issuing occasional requests for proposals in areas or on specific topics where Bank research is weak.
- The approval process for all research needs a better balance of academic quality and policy relevance than is currently the case. While external reviews are currently obtained by the Research Committee, the large fraction of research that does not go through the Research Committee also needs to be subject to this kind of review.
- The Bank needs to ensure that more of the senior managers of its central research arm, the Development Economics group, are as well-qualified in statistical and

econometric methods as they are in economics. Ideally, this recommendation should apply throughout the Bank, but DEC is the obvious home for this expertise and for quality control, through the Chief Economist, or perhaps through a “Chief Statistician” in DECDG.

Data

- The Bank needs to review its data collection, archival, and dissemination procedures. The Development Economics Data Group needs to be brought into closer contact with researchers, and should become a center of statistical advice on survey collection, as well as taking the lead on archiving and dissemination of all Bank data. This would involve the strengthening of its statistical and econometric expertise.
- The Bank should take a stronger lead in promoting the international harmonization and dissemination of household surveys.
- As the Bank’s original data collection function expands, it needs to put into place standard protocols for the release and revision of data. We also recommend a regular (internal and external) review of data collection priorities.

Quality control and flagships

- The Bank should consider ways of enhancing the role of the Chief Economist (as Chief Economist, rather than through DEC) in quality control over all Bank publications, particularly flagships. The effectiveness of flagships in disseminating best practices needs a thorough review with an emphasis on maintaining consistent high quality rather than quantity. We suspect that the

plethora of flagships strains the Bank research leadership's capacity to monitor quality and reliability. Flagships are also an area where the Bank must be especially careful to present a balanced picture of research results and debate among serious policy researchers.

Relationships with academia

- The Bank should foster closer relationships with academic researchers, including better visiting programs, and through conferences.
- The research function of the ABCDE conference should be reviewed.

Capacity building

- The Bank should make greater efforts to foster collaborative work between Bank and developing country researchers, possibly through greater institutional support in the countries. In any case, it should aim to increase the representation of developing country researchers in its research output.

wb18913

L:\DECERS\Evaluation of Research\Final Version\RESEARCH EVALUATION 2006 Main Report.doc

11/20/2006 3:43:00 PM