

If teachers were paid more based on their performance, would they be more effective? Would teachers teach better and would students learn more? Can teacher pay-for-performance (PFP) programs motivate better performance?

Do teacher PFP programs work? Is teacher performance pay the right approach for a given country? This note presents evidence about the effectiveness of teacher PFP programs and outlines criteria for considering teacher performance pay in a given country context.



# Teacher Pay-for-Performance: What Works? Where? And How?

Mary Breeding, Tara Bêteille and David K Evans

APRIL 2021

Financial incentives are one way of motivating teachers to perform better by providing them with rewards based on direct measures of their performance. Research evidence from high-income and low-income countries suggests that traditional ways of rewarding teachers based on their seniority or level of education are often ineffective in improving student learning.<sup>i</sup> Teacher PFP programs represent an alternative approach for motivating better teacher performance. Teacher PFP can be attractive to policymakers because PFP does not involve making fundamental changes to teacher pay-scales, unlike many career-based incentives, and these programs can be suspended more easily if ineffective.<sup>ii</sup> The evidence on their effectiveness, however, is mixed, with promising results in some education systems such as in India (Andhra Pradesh), Chile, Rwanda, and Tanzania, but either negative or inconclusive results in other cases.<sup>iii</sup> Few PFP programs have been sustained over time in low- and middle-income countries.

Teacher PFP programs represent an alternative approach for motivating better teacher performance. Teacher PFP can be attractive to policymakers because PFP does not involve making fundamental changes to teacher pay-scales, unlike many career-based incentives, and these programs can be suspended more easily if ineffective.

## Box 1. Motivating Teacher Performance: Different Extrinsic Mechanisms<sup>xi</sup>



**Career-based incentives** provide teachers with opportunities for promotion or pension increases and are typically linked to job tenure.



**Accountability systems** seek to identify and improve the performance of weak teachers and/or remove poor-performing teachers. These may include a mix of carrot approaches (that reward good performance) and stick approaches (that take punitive measures for poor performance).



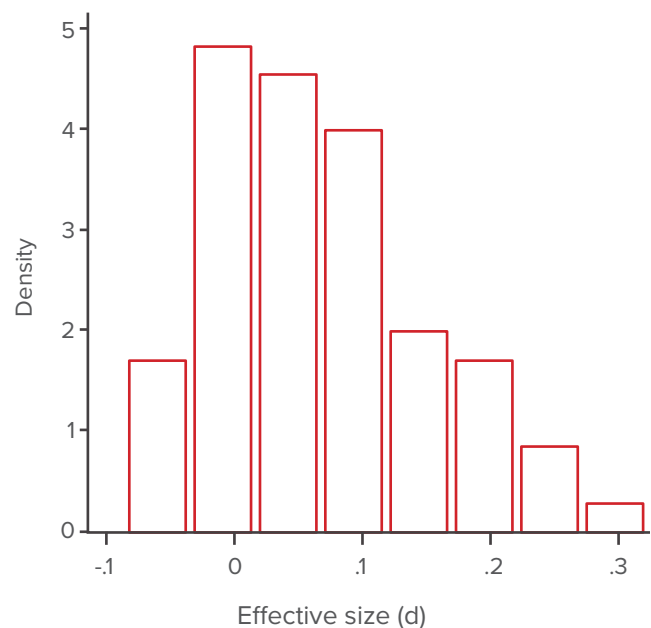
**Financial Incentives** include teacher performance pay and salary increases for performance. PFP programs provide financial incentives—monetary bonuses or in-kind gifts, above standard teacher salaries—as a direct reward to teachers for improved performance. These may be based on measures of individual teacher performance or group-based performance, and they are provided to improve teacher performance and recruitment, or to retain high quality teachers (Bruns and Luque 2015).

Across programs (72 reported outcomes of impacts on student test scores in 15 evaluations), effect sizes range in magnitude from a minimum effect of -0.08 SD and a maximum effect of 0.32 SD increase in student test scores. The median reported effect size is a 0.06 SD increase.

### Does Teacher Performance Pay Work?

Evidence about the effectiveness of teacher PFP programs is mixed, particularly in low and middle-income countries.<sup>iv</sup> A review of teacher PFP programs by Breeding, Béteille and Evans (2021) examines the evidence of impact that teacher PFP programs have had on student learning and teacher outcomes in low and middle-income countries.<sup>v</sup> The authors develop a framework for reviewing teacher PFP programs and apply this framework to a sample of all rigorously evaluated teacher PFP programs in developing countries (15 in total), highlighting trends associated with effective teacher PFP. Effects on student learning outcomes vary substantially. Across programs (72 reported outcomes of impacts on student test scores in 15 evaluations), effect sizes range in magnitude with a minimum effect of -0.08 Standard Deviations (SD) and a maximum

**Figure 1: Only a few Performance Pay Programs have improved student learning – most have not**



Sample: 72 reported outcomes of impacts from 15 evaluations

effect of 0.32 SD increase in student test scores. The median reported effect size is a 0.06 SD increase. Most programs produce a mix of outcomes across different measures of test scores and subjects, most of which are substantively and statistically insignificant when it comes to increasing student test scores.<sup>vi</sup>

# What can we learn from successful PFP programs?

Breeding, Béteille and Evans (2021) explore shared characteristics among successful programs and highlight three design and implementation features associated with PFP programs that successfully improve student performance and teacher outcomes.

## Successful PFP Programs:



**1. Match the right reward size and kinds of incentives (individual vs. group-based incentives) to the country context;**



**2. Ensure accountability with multiple monitoring and evaluation measures (not just test scores); and**



**3. Have built-in considerations for program sustainability beyond a given period of evaluation.<sup>vii</sup>**

In addition, two considerations, not easily quantified, but that are highlighted by evaluators as important considerations for PFP programs include the importance of teacher support and understanding as well as the role of governments, especially at the district and local level. For instance, when government capacity is particularly low the ability to implement the program may hinder potential success. Even among the most well-designed evaluations, capacity of school, district and government counterparts to implement programs is required for successful program execution.

First, programs that produce positive outcomes in student test scores have developed well-researched incentives that teachers are likely to respond to based on the local context and types of teachers targeted. Benefits are either provided at the school-level or to individual teachers based on perceived responsiveness in these systems and may be distributed as cash or in-kind gifts depending on what is culturally appropriate. For example, one program in Tanzania used cell phones as an effective intervention after consultation.<sup>viii</sup> Second, successful PFP programs ensure multiple mechanisms for accountability. The evaluated PFP programs with successful outcomes all include multiple methodologies for monitoring teachers, ranging from classroom observation to surveys of key actors. For instance, both the Rwanda Education Board PFP Program and the Kenya Incentive Program included classroom observation, teacher surveys and surveys of head teachers, and the Rwanda PFP program conducted additional “Lab-in-the-field” experiments to capture additional attributes of teachers.<sup>ix</sup> A final characteristic of successful programs is the consideration for program sustainability. The Sistema Nacional de Evaluación del Desempeño de los Establecimientos Educacionales (SNED) program in Chile was the only wide-spread government program implemented as an institutionalized policy among the programs from low- or middle-income countries included in this review that produced positive and statistically significant outcomes. (Chile is a high-income country now, but it was a middle-income country when the program was introduced and evaluated.) The SNED built in plans for sustainability of the program from the design stage, starting with the development of the government-adopted policy. Further, the SNED incorporated mechanisms for ongoing consultation with teachers, and the policy was adopted alongside a long-term set of accountability reforms to teacher evaluations and student testing.<sup>x</sup> Table 1 highlights specific lessons learned from evaluations that show positive impacts on student learning outcomes.

**Table 1. Lessons from Evaluations of Effective Teacher PFP Programs**

PROGRAM NAME	INTERVENTION	OUTCOME	KEY LESSON(S)
<b>Andhra Pradesh Randomized Evaluation Study (AP RESt—India<sup>xii</sup></b>	Treatment 1: Individual incentives; Treatment 2: Group-based Incentives;	Teacher bonuses improved teacher performance and student test scores in both treatment groups: At the end of two years of the program, students in incentive schools performed significantly better than those in control schools by 0.28 and 0.16 standard deviations in math and language tests respectively.	Individual incentives resulted in higher gains in student learning than group-based incentives.
<b>Kenya Teacher Incentive Program<sup>xiii</sup></b>	Treatment 1: Group-based In-kind prizes (seen as more socially acceptable than cash in Kenya) received based on average performance of children in the school; Control Group: No support	Student test scores increased on exams linked to the incentives, but not on other, unrelated exams. Teacher attendance and homework assignments were unaffected, but the number of test preparation sessions for students increased.	Weak sustainability: Student test scores improved slightly in the first 2 years, and then returned to pre-program results after program ended.

PROGRAM NAME	INTERVENTION	OUTCOME	KEY LESSON(S)
<b>Twaweza experiment—Tanzania</b> <sup>xiv</sup>	Treatment 1: Individual teacher incentives; Treatment 2: Extra resources through capitation grants; Treatment 3: Combined individual teacher incentives with capitation grants, Control group: No support	Extra resources had no impact on learning outcomes. Teacher incentives alone had some positive impact. Teacher incentives with extra resources had stronger impacts on learning outcomes.	Incentives improved learning outcomes most effectively when coupled with extra resources.
<b>Rwanda Education Board—PFP Program</b> <sup>xv</sup>	Treatment 1: Fixed Wage end-of-year payout roughly RWF 20,000 (3% of teachers annual salary in imihigo system); Treatment 2: PFP end-of-year payout of RWF 100,000 for those in the top quintile, or zero otherwise.	PFP changed the composition of the teaching workforce by attracting more money-oriented job candidates, and PFP contracts improved teacher effort. The total estimated effect of PFP to on student learning was 0.20 SD.	PFP can improve teacher effort and recruiting for PFP contracts does not change the distribution of teacher skills.
<b>National System of School Performance Assessment (NSPED)-Chile</b> <sup>xvi</sup>	Treatment 1: Opportunity for individual teachers to win a monetary productivity bonus. Control: No support	Student test scores increased by 0.14-0.25 SD on language and math tests for school tournament winners.	Solicit teacher and community support at each stage of program implementation for effective results; Incentives worked with strong accountability systems in place.

## Is Teacher Performance Pay the Right Approach for My Country?

PFP programs are a costly and sometimes politically contentious policy option, which usually involves rewarding high-performing teachers with additional pay increases. This section provides a step-by-step decision guide based on research to help policymakers decide whether PFP programs are likely to work in their country. Three evidence-based considerations for examining possible teacher PFP programs include: (1) Reviewing necessary pre-conditions, (2) reviewing important design and implementation features, and (3) evaluating risks that could derail a teacher PFP program.

**Step 1. Are the right preconditions in place?** Technical requirements, resources, and political will are three necessary (albeit not sufficient) pre-conditions for a successful teacher PFP program. Technical requirements include having a sufficient accountability system with a capable bureaucracy and data systems.<sup>xvii</sup> In order to effectively implement a teacher PFP program, systems need capacity to reliably identify high-performing and poor-performing teachers. In addition to rewarding high-performers, systems in high-income countries that show promise for improving student performance

typically incorporate mechanisms for both (1) working to improve the quality of poor performing teachers (e.g., through targeted teacher professional development); and (2) for removing poor performers. For instance, in Singapore, the Enhanced Performance Management System provides professional development and counseling opportunities to the least effective five percent of teachers to help them improve their performance.<sup>xviii</sup> Similarly, an evaluation and feedback system for teachers in Washington, DC (USA) provided an avenue for poor performing teachers to voluntarily leave (thereby improving the overall stock of existing teachers), and how DC policymakers instated a one-year probationary period with mandatory coaching for poor-performers to improve their performance ratings, after which poor-performers are subject to dismissal.<sup>xix</sup>

Additionally, ensuring sustained resources and political support for teacher PFP is critical. Pay increases are costly to provide. Across evaluated teacher PFP programs in low and middle-income countries, the 25th percentile is two percent of teachers' annual salaries, and the 75th percentile is 8.25 percent of teachers' annual salaries.<sup>xx</sup> Teacher PFP

**Technical requirements, resources, and political will are three necessary (albeit not sufficient) pre-conditions for a successful teacher PFP program.**

## Across evaluated teacher PFP programs in low and middle-income countries, the 25th percentile is two percent of teachers' annual salaries, and the 75th percentile is 8.25 percent of teachers' annual salaries.

programs require numerous resources from the program design stage to carrying out actual incentive payments, and to the accountability frameworks for monitoring and evaluating program implementation.

The importance of political will and teacher buy-in are important considerations for the design and implementation of teacher PFP programs. Questions arise about how to design programs that will be supported by teachers, teachers' unions, and various levels of school administration. The 2018 World Development Report as well as Bruns and Luque (2014) discuss the Chile example as a case study emphasizing the importance of gaining teacher support at every stage of the development of Chile's incentive and teacher accountability program.<sup>xxi</sup> In particular, one way the government was able to secure buy-in of teachers was to ensure that schools in areas with difficult geographic or socioeconomic conditions were able to compete on an equal footing with more advantaged schools by

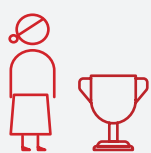
stratifying schools into homogenous groups. In addition to incorporating teachers' views from the outset and working to ensure equity across schools, Chile included an initial period for teachers to opt-in/out of the program (as have other successful programs such as the IMPACT program in Washington, DC), in which initial program participation was voluntary. This approach seems to have helped emphasize the potential rewards of these systems and to have helped secure support of teachers who may have otherwise been skeptical of participating in the PFP program.<sup>xxii</sup>

Similarly, in the case of Andhra Pradesh, the evaluators stress the importance of having teacher support by framing performance pay programs appropriately. Drawing on evidence from psychology, the authors discuss the importance of framing incentive programs in a manner that intrinsically motivates teachers by recognizing the difficult job teachers do and the importance of good teaching, since extrinsic incentives can be viewed as a way of seeking to exercise control over workers and can actually lead them to feel less motivated.<sup>xxiii</sup>

### Step 2. What design and implementation features should be considered when developing teacher PFP program?

Lessons from existing teacher PFP programs point to four essential design questions related to teacher PFP programs: Who gets rewarded; what gets rewarded; how rewards are provided; and the form and size of rewards.

## Box 2. Teacher Performance Pay Design Features<sup>xxiv</sup>



### Who gets rewarded?

PFP programs target individual teachers or provide group-based, school-wide benefits to teachers based on specific measures of performance. In addition to teachers, school-level benefits may also target school principals and administrators. There is some evidence suggesting that individual benefits may be more effective than group-based benefits (Muralidharan and Sundararaman 2011), but there is no clear consensus.



### What gets rewarded?

Objectives of teacher PFP programs include rewarding teacher effort, teacher value-added, student achievement, student participation, teaching and learning frameworks, teacher's commitment to school and/or community, or other measures. Evidence suggests that having more than one measure of performance may yield more valid and reliable results in acknowledging and rewarding high-performing teachers, and that strong systems of accountability help to improve these measures.



### How do rewards get distributed?

The timing and mechanism for reward distribution may influence the degree of reward effectiveness. For instance, teachers may behave differently when rewards are based on a loss-aversion model, where bonuses were framed as losses toward the end of the year, relative to a traditional model of bonus pay (gains) at the end of an academic year. In an experiment in Chicago public schools, Fryer et al. (2012) find that when bonuses were framed as potential losses, student math test scores improved significantly by measures of 0.2 and 0.4 SD.



### What form to give rewards in and how much?

Bonuses can take the form of monetary teacher bonuses or in-kind rewards. The value of these bonuses can vary substantially, but there is evidence that the size of the bonus matters to the context and that it needs to be enough to motivate teacher effort (Bau and Das 2017). In a recent evaluation assessing the impact of a pay-for-percentile system vs. rewarding teachers based on certain proficiency thresholds in Tanzania, the evaluators find that rewarding proficiency thresholds is equally effective at improving student test scores.

**Step 3: What could derail performance pay interventions?** Well-designed PFP programs evaluate risks and incorporate mechanisms into the program design to help curtail sub-optimal effects related to implementation.<sup>xxv</sup> Commonly reported negative effects of PFP programs include cheating<sup>xxvi</sup>, “teaching to the test” with an emphasis on producing short-term results from students<sup>xxvii</sup>, or test manipulation by either excluding weak students or teaching to a specific group of students who show promise of better performance.<sup>xxviii</sup> Free-riding and moral-hazard problems can further occur in the provision of group-based incentives.<sup>xxix</sup> To help reduce the risks, it is important to ensure that standardized student achievement tests have multiple checks for accuracy and that accountability systems are in place to routinely monitor teacher classroom behavior and performance.

## Teacher PFP Programs in Low and Middle-Income Countries—At a Glance

### What do Teacher Performance Pay Programs Look Like in Low- and Middle-Income Countries?



**Types of Incentives:** Out of 15 evaluated programs in low and middle-income countries, 9 provided individual monetary incentives, 5 provide group-based monetary incentives, and 1 provided in-kind group-based incentives. One program, provided both individual and group-based monetary incentives.

» Evidence from the one study to contrast individual and group incentives suggests that individual incentives may be slightly more effective than group based.



**Incentive Size:** Across evaluated teacher PFP programs in low and middle-income countries, the 25th percentile is two percent of teachers’ annual salaries, and the 75th percentile is 8.25 percent of teachers’ annual salaries. Taken together, the papers do not provide adequate information to identify a link between incentive size and program success.



**Program Age:** The average program age at the time of evaluation was 4 years (Min 1 year, Max 14 years).



**Suboptimal Effects:** 20 percent (3 out of 15 evaluations) report sub-optimal effects such as cheating or teachers teaching to the test during program implementation.



**On sustainability,** 33 percent of evaluated teacher PFPs in low and middle-income countries are connected to broadly adopted government policies—sustainable beyond the evaluation period (Indonesia, Zambia, Chile, Uruguay, and Punjab, Pakistan). Of these, only Chile has seen positive and significant outcomes on student achievement.

# Endnotes

<sup>i</sup> Hanushek (2006).

<sup>ii</sup> Bruns and Luque (2015).

<sup>iii</sup> Contreras and Rau (2012); Lever and others (2019); Mbiti and others (2017); Muralidharan & Sundararaman (2011).

<sup>iv</sup> Teacher performance pay programs have been implemented in many more high-income countries, and results in high-income countries are equally mixed. Woessman et al. (2011) find that PISA scores in countries with performance pay systems are approx. 0.25 SD higher than those without. However, in reviews of impact evaluations, there is little evidence that performance pay improves student achievement. Only four out of 28 papers from high-income countries reviewed for this analysis find successful teacher performance pay programs that resulted in improved student learning outcomes.

<sup>v</sup> Breeding, Beteille, and Evans (2019).

<sup>vi</sup> These results align with a meta-analysis by McEwan (2015), in which he finds the average effect size of both teacher and student incentives (combined) is 0.09 SD. Further, in his meta-analysis looking at how different types of education interventions influence student achievement scores, he finds that the effect of performance incentives on student achievement tests in meta-regressions is a 0.056\*\* SD increase (full model controlling for implementers, region, and other characteristics of the program) and 0.042 (not significant, partial model). Both reviews of teacher PFP programs highlight the range of outcomes associated with PFP programs, and they raise questions about the real feasibility of these programs if such programs given their limited effects on student learning.

<sup>vii</sup> Breeding, Beteille, and Evans (2019).

<sup>viii</sup> Sabarwahl and Others (2019).

<sup>ix</sup> Glewwe and others (2010); Leaver and others (2019).

<sup>x</sup> Contreras and Rau (2012).

<sup>xi</sup> Bruns and Luque (2015). Behrman and others (2012); Duflo and others (2012); Mbiti and others (2017).

<sup>xii</sup> Muralidharan and Sundararaman (2011).

<sup>xiii</sup> Glewwe and others (2010).

<sup>xiv</sup> Mbiti and others (2017).

<sup>xv</sup> Leaver and Others (2019).

<sup>xvi</sup> Contreras and Rau (2012).

<sup>xvii</sup> Dee and Wyckoff (2015); Bruns and Luque (2015)

<sup>xviii</sup> Tucker (2011).

<sup>xix</sup> Dee and Wyckoff (2015)

<sup>xx</sup> Breeding, Beteille, and Evans (2019).

<sup>xxi</sup> World Bank (2017); Bruns and Luque (2015).

<sup>xxii</sup> Dee and Wyckoff (2015).

<sup>xxiii</sup> Muralidharan and Sundararaman (2011); Fehr and Falk (2002) argue that designers of incentives need to consider human emotions, specifically the desire to reciprocate and the desire to avoid social disapproval. They provide experimental evidence that poorly devised incentives can actually backfire stimulating incentives for retaliation through strikes or protest.

<sup>xxiv</sup> Muralidharan & Sundararaman (2011). Mbiti and others (2019); Freyer and others (2012); Bau and Das (2017).

<sup>xxv</sup> Ibid.

<sup>xxvi</sup> Behrman and others (2012); Muralidharan and Sundararaman (2011).

<sup>xxvii</sup> Holmstrom and Milgrom (1991).

<sup>xxviii</sup> Jacob (2005); Neal and Schanzenbach (2008).

<sup>xxix</sup> Glewwe and others (2010).

# References

- Bau, N. & J. Das (2017). "The misallocation of pay and productivity in the public sector: Evidence from the labor market for teachers." Policy Research working paper; no. WPS 8050. Washington, D.C.: World Bank Group.
- Breeding, M., T. Beteille, & D.K. Evans. (2019). "Teacher Pay-for-Performance (PFP) Systems: What works? Where? And how?" Teachers Thematic Group Working Paper. Washington, DC: World Bank.
- Behrman, J., S. Parker, P. Todd, & K. Wolpin. (2015). "Aligning learning incentives of students and teachers: Results from a social experiment in Mexican high schools." *Journal of Political Economy*, 123(2): 325-364.
- Bruns, B. & J. Luque. (2015). *Great teachers: How to raise student learning in Latin America and the Caribbean*. Washington, DC: World Bank.
- Contreras, D., & T. Rau. (2012). "Tournament incentives for teachers: Evidence from a scaled-up intervention in Chile." *Economic Development and Cultural Change* 91 (1): 219–46
- Dee, T. & J. Wyckoff. (2015). "Incentives, selection, and teacher performance: Evidence from IMPACT." *Journal of Policy Analysis and Management* 34(2): 267-297.
- Fehr, E. & A. Falk. (2002). "Psychological foundations of incentives." *European Economic Review*, 46, 687-724.
- Fryer, R.G. (2013). "Teacher incentives and student achievement: Evidence from New York City public schools." *Journal of Labor Economics* 31(2).
- Glewwe, P. W., Nauman I., & M.R. Kremer. (2010). "Teacher incentives." *American Economic Journal: Applied Economics* 2 (3): 205–27.
- Hanushek, E. (2006). "School Resources," in *Handbook of the Economics of Education (Vol 2)*, ed. by E. Hanushek, and F. Welch: North-Holland.
- Holstrom, B. & P. Milgrom. (1987). "Aggregation and linearity in the provision of intertemporal incentives," *Econometrica*, 55, 303-28.
- Jacob B.A. (2005). "Accountability, Incentives and Behavior: The Impact of High-Stakes Testing in the Chicago Public Schools," *Journal of Public Economics*, 89, 761-96
- Leaver, C., O. Ozier, P. Serneels, & A. Zeitlan. (2019). "Recruitment, effort, and retention effects of performance contracts for civil servants: Experimental evidence from Rwandan primary schools." Draft Results.
- Mbiti, I., Muralidharan, K., Romero, M., Schipper, Y., Manda, C., & Rajani, R. (2019). "Inputs, incentives, and complementarities in education: Experimental evidence from Tanzania." *The Quarterly Journal of Economics*, 134(3), 1627-1673.
- McEwan, Patrick J. (2015). "Improving learning in primary schools of developing countries: A meta-analysis of randomized experiments." *Review of Educational Research* 85(3), 353-394.
- Muralidharan, K., & V. Sundararaman. (2011) "Teacher Performance Pay: Experimental Evidence from India." *Journal of Political Economy* 119(1), 39-77.
- Neal, D. & D. W. Schanzenbach. (2008) "Left Behind by Design: Proficiency Counts and Test-Based Accountability," University of Chicago.
- Sabarwal, S. and M. Abu-Jawdeh. (2018). "What Teachers Believe: Mental Models About Accountability, Absenteeism, and Student Learning." Policy Research Working Paper; no. 8454; Washington, DC: World Bank.
- Sabarwal, S., Filmer, D. & J. Habyarimana. (2019). "What Should We Expect from Incentives?" Washington, DC: RISE 2019 Conference presentation.
- Tucker, M.S. (2011). *Surpassing Shanghai: An Agenda for American Education Built on the World's Leading Systems*. Cambridge, MA: Harvard Education Press
- Woessman, L. (2011). "Cross-Country Evidence on Teacher Performance Pay." *Economics of Education* 30(3), 404-418.
- World Bank. (2017). *Learning to Realize Education's Promise*, World Development Report. Washington, DC: World Bank.