RESULTS-BASED FINANCING IN EDUCATION

# Learning from What Works

WORLD BANK GROUP
Education

Education Global Practice
Results in Education for All Children (REACH)

## Acknowledgements

*This report was written by Jessica D. Lee and Octavio Medina. We are grateful to Omar Arias, Practice Manager, and Samer Al-Samarrai, Senior Economist and Results in Education for All Children (REACH) program manager, for their guidance and support. Helpful comments and feedback were also provided by Mariam Adil, Rabia Ali, Juan Baron, Deon Filmer, Andrea Guedes, Alaka Holla, Peter A. Holland, Sarah Holzapfel, Alice Kunz, and Shwetlena Sabarwal. We also want to thank all of the staff from various development agencies who took the time to speak with us to share their knowledge.*

*The REACH trust fund is supported by the governments of Germany, Norway, and the United States.*

*The findings, interpretations, and conclusions expressed in this paper are entirely those of the authors. They do not necessarily represent the views of the International Bank for Reconstruction and Development/World Bank and its affiliated organizations, or those of the Executive Directors of the World Bank or the governments that they represent.*

# Contents

**Acronyms** ..... 6

## Executive Summary

**Incentives to Teachers, Students and Families, and Schools** ..... 8
**Incentives to Governments** ..... 8

## Introduction

**Which Definition of RBF is Being Used?** ..... 10
**Why Focus on RBF in Education?** ..... 10
**Why Examine Four Different Levels?** ..... 11
**Methodology** ..... 12
- Methodology to Collect Operational and Tacit Knowledge ..... 12
- Survey Design ..... 14

**Results-based Financing in Education: Donor Portfolios**
- ADB (Asian Development Bank) ..... 16
- AFDB (African Development Bank) ..... 16
- DFID (Department for International Development) ..... 16
- GPE (Global Partnership for Education) ..... 16
- IDB (Inter-American Development Bank) ..... 16
- NORAD (Norwegian Aid Agency) ..... 17
- SIDA (Swedish International Development Cooperation Agency) ..... 17

## Part I – When Do Incentives Work?

**Design Issues** ..... 19
**Using RBF with Teachers** ..... 19
- The Theory Behind Using RBF with Teachers ..... 19
- Does Using RBF with Teachers Improve Outcomes? ..... 20
- Design Issues: What to Incentivize (Outputs or Outcomes?) ..... 22
- Design Issues: Who to incentivize? ..... 22
  - *Individual or Group-based* ..... 22

Design Issues: Which Metrics to Use ..... 23
*What Metric to Use: Level, Piece Rate, or Rank* ..... 23
Design Issues: Behavioral Responses ..... 24
Design Issues: Sustainability ..... 25
Design issues: Long term effects ..... 25
Design Issues: Gaming and Cheating ..... 25
Conclusion ..... 26
**Using RBF with Students and Families** ..... 27
The Theory behind Using RBF with Students and Families ..... 27
Does Using RBF with Students and Families Improve Outcomes? ..... 27
Design Issues: The Role of Conditionality ..... 30
Design Issues: The Role of Information and Labeling ..... 30
Design Issues: What to Incentivize ..... 31
Design Issues: Who Should Be Incentivized? ..... 32
Conclusion ..... 33
**Using RBF with Schools** ..... 33
The Theory behind Using RBF with Schools ..... 33
Does Using RBF with Schools Improve Outcomes? ..... 34
Design Issues: Competitive Distribution of Resources ..... 35
Design Issues: Equity ..... 36
Design Issues: Household response ..... 37
Design Issues: Long term effects ..... 37
Conclusion ..... 37
**Combining RBF Interventions to Overcome Constraints** ..... 38
Targeting different levels ..... 38
Combining RBF and institutional capacity building ..... 38
**Summary** ..... 40

## Part II – RBF and Governments: Making RBF More Effective

Implementation ..... 41
Planning ..... 41
Design ..... 41

**Choosing RBF: Commitment, Cautions, Cost, Context** ..... 43
Commitment ..... 44
*Recipient Feels "Pressured" into Accepting RBF* ..... 45
*Recipient is Keen to Pursue RBF and Has the Necessary Political Will* ..... 45
Costs and Benefits (Advantages and Disadvantages of RBF Over Traditional Aid) ..... 46
*Cost-effectiveness of RBF* ..... 47
Cautions ..... 49
Context ..... 50
*Country Systems* ..... 50
*Capacity* ..... 51
*Conflict, Fragility, and Violence* ..... 53
**Design Priorities** ..... 54
Cascading Incentives ..... 55
Selecting and Pricing Indicators ..... 57
*DLI Analysis: The Basics* ..... 57
*DLIs and Results Chains: Few DLIs Focus on Outcomes* ..... 59
*Examples of Results Chains: Bangladesh, Lebanon, and Tanzania* ..... 62
*DLIs and Short Project Timelines* ..... 68
*Pricing DLIs: Three Hypotheses and a Heuristic* ..... 69
*DLIs: Scalability and Disbursement Models* ..... 72
*Zero/Global DLIs* ..... 73
**Adaptive Implementation** ..... 74
Monitoring and Information Systems ..... 74
*Purpose of Education Management Information Systems (EMIS)* ..... 74
*Building Systems for RBF* ..... 75
*Level of Complexity Needed for Systems* ..... 75
*Monitoring Options* ..... 76
Verification ..... 77
Gaming and Cheating ..... 80
Implementation Quality ..... 81
**Failure to Achieve Targets** ..... 84
**Sustainability** ..... 85

**Concluding Remarks** ..... 86

**References** ..... 88

## Acronyms

| | |
|---|---|
| African Development Bank | AFDB |
| Asian Development Bank | ADB |
| Conditional Cash Transfers | CCT |
| Department for International Development | DFID |
| Disbursement-Linked Indicators | DLI |
| Education Management Information System | EMIS |
| Financial Management | FM |
| Fragile and Conflict/Violence affected | FCV |
| Girls Education Challenge | GEC |
| Global Partnership for Education | GPE |
| Health Results and Innovation Trust Fund | HRITF |
| Information Technology | IT |
| Inter-American Development Bank | IADB |
| International Anti-Corruption Resource Center | IACRC |
| Investment Project Financing | IPF |
| Management Information System | MIS |
| Management System for Education Quality | SIGCE |
| Non-Governmental Organization | NGO |
| Norwegian Aid Agency | NORAD |
| Program for Results | PforR |
| Programa de Asignación Familiar | PRAF |
| Project Appraisal Document | PAD |
| Public Financial Management | PFM |
| Quality Assurance System | QAS |
| Results in Education for All Children | REACH |
| Results-Based Financing | RBF |
| Sistema de Administração Financeira do Estado | SISTAFE |
| Standard Deviation | SD |
| Swedish International Development Cooperation Agency | SIDA |
| The Vaccine Alliance | GAVI |
| United States of America | US |

# Executive Summary

**Results-based financing (RBF) has gained popularity in the international development community because of its potential to make education spending more effective and efficient.** In the education sector, RBF has been primarily applied to four levels: teachers; students and families; schools; and governments. The results overall have been mixed, with some notable successes and some disappointing experiences.

**This report explores when and how RBF can help achieve better impacts in education.** While there is no rigorous evidence available to suggest that RBF on its own is better at producing learning outcomes relative to other development financing modalities, there is a significant amount of research that shows RBF can have positive effects by incentivizing specific stakeholders in the education system.

In addition, there is operational evidence available on how RBF can be designed and implemented with country partners more effectively. It is important for practitioners and policymakers to learn from this evidence as the RBF portfolio in education grows across development agencies.

## Incentives to Teachers, Students and Families, and Schools

Though the research is neither comprehensive nor definitive, there is substantial evidence to suggest the following:

1. **RBF and teachers:** Teacher incentives can but do not always improve teacher attendance and student learning. The design of the incentive scheme and the context matter. The effects are larger and more positive in developing country contexts.
2. **RBF and students and families:** Student and family incentives (such as CCTs, for instance) has a good track record of reducing school dropout and increasing school attendance, though the evidence for its effects on student learning is more mixed. Conditional transfers to students tied to their own learning are a promising area of future research.
3. **RBF and schools:** The evidence base on the effectiveness of performance-based grants is still quite limited. For now, it seems that in some cases they can work, especially when grants are combined with other interventions such as capacity building (for example, to principals and school committees) or when money is spent on inputs that affect learning outcomes.
4. **Synergies:** There is growing evidence that combining different RBF interventions within the same program can generate results that go beyond the sum of any two interventions alone. Though the research is limited, this suggests that RBF that tackles several bottlenecks at once can have larger effects.

## Incentives to Governments

There is much less robust research available on results-based financing arrangements between donors and country governments. Though there is some evidence from other sectors such as health, few of these programs have been rigorously evaluated in education. However, there is a large base of operational knowledge across multiple agencies, which points at several key criteria for more effective RBF:

1. Choosing RBF as the appropriate financing modality requires careful consideration of political commitment, and understanding the risks involved, costs, and country context (for example, capacity and country systems).
2. RBF project design should prioritize the cascading of incentives and should select and price indicators with an objective or methodology in mind. Some of these include cost effectiveness, increasing the chances of achieving other indicators, or reducing risk of nonpayment.
3. RBF project implementation should think of the purpose of monitoring and information systems, invest upfront in verification, and be adaptive and flexible in order to address realities on-the-ground.

# Introduction

It is not clear that the majority of development financing has been either effective or sustainable, and many stakeholders in international development are keen to change this. In recent years, **results-based financing (RBF) has been championed as a way to increase both the efficiency and effectiveness of aid.** While there are those who defend it and others who decry it, almost everyone wants to know: **does it work?**

For various reasons that will be explored in this report, this is not a yes or no question. However, it is possible to investigate when and how RBF can improve results in education.

**This report will outline the theories, design considerations, implementation issues, and impact of RBF in the education sector** and examine how RBF has worked, or not worked, when used with teachers, students and families, schools, and governments.

Specifically, this report will examine three types of evidence: (i) particular types of RBF interventions, such as teacher incentives, on which many studies have been done; (ii) the operational knowledge of development agency staff who have designed and implemented RBF projects in education; and (iii) documentation from RBF projects from different stages of the project cycle.

Many of the lessons can be generalized across sectors, though we will draw our examples primarily from the education and health sectors.

## Which Definition of RBF is Being Used?

The world of results-based financing is populated by an alphabet soup of acronyms. For the purposes of this report, results-based financing (RBF) is an umbrella term referring to any program or intervention that provides rewards after the credible verification of an achieved result. These rewards can be monetary or non-monetary and can be partial (such as a bonus on top of a salary) or whole (such as the cost of training a teacher under output-based aid).[1]

There are differing opinions on what actually constitutes results-based financing, with much of the debate centered on what constitutes a "result." In this report, results are defined broadly. They can be outputs (such as the implementation of a new teacher training system), intermediate outcomes, final outcomes (such as learning) or — more likely — a mixture. Importantly, the dividing line between inputs and outputs can depend on which particular bottleneck the RBF is being used to resolve and on the objectives of each specific project.[2]

## Why Focus on RBF in Education?

There has been less research, either qualitative and quantitative, done on RBF in education than other social sectors, such as health. There are various reasons for this, including the fact that there have been more long-standing examples of RBF in the health sector, notably through the World Bank-managed Health Results and Innovation Trust Fund (HRITF) and the public-private partnership GAVI (the Vaccine Alliance), both of which have specific goals and target populations. In the original GAVI scheme, bonuses were paid out to every additional child over the baseline who received vaccinations.[3] Many of the early HRITF interventions were aimed at increasing the use of maternal and neonatal care services, which translated into indicators such as immunizations, clinic visits, and the delivery of babies in a health facility.[4] These types of indicators are inherently quantitative and thus more easily measured than outcomes such as learning.

The World Bank's 2018 World Development Report states that, even when learning is the explicit goal, achieving that goal can be difficult because tasks within the education system are often carried out in a fragmented way by many different actors, which dissipates accountability.[5] This can make it challenging to accurately identify the binding constraints in any particular country's education system, much less know which constraints can be overcome by incentives. Moreover, education indicators are often not inherently quantitative, particularly when related to quality. For example, simply training teachers does not necessarily lead to better learning outcomes, nor does increasing enrollment rates. The education system and theories of change within the system are complex and contain multiple actors whose actions must be aligned in order for learning to occur.[6] Thus, it is important to assess the promises and pitfalls of RBF to help different stakeholders in the education sector to understand what it can and cannot accomplish.

1 World Bank (2017)
2 World Bank (2017)
3 Pearson et al (2010)
4 World Bank (2015)
5 World Bank (2017c)
6 World Bank (2017c)

## Why Examine Four Different Levels?

This report focuses on four different levels: teachers, students and families, schools, and governments. Why look at different levels of RBF interventions? Simply put, the education sector in every country is not a monolithic entity, but a system with many moving parts.[7] The role of RBF interventions varies based on the level of the intervention, and so do the relevant actors involved in the process.

Of course, some variables are common to all levels. An actor or an institution's performance depends on variables like motivation, inputs, and skill sets. But these can take different forms. For example, at the teacher level, a constraint may be whether a teacher shows up or has the content and pedagogical knowledge to teach. At the student and family level, a constraint is getting the child to attend school (and then learn). At the school level, managing inputs effectively becomes crucial, whereas at the government level policy and incentive design may take precedence. The problems and variables are related, but the constraints and possible solutions vary.

As such, while there are things in common, the guidance and design issues can be quite different. This analysis at different levels can provide more nuance than treating RBF interventions in education as a homogeneous whole.

*Table 1: Levels, Roles and Constraints in Education Systems*

| LEVEL | ROLE | SAMPLE CONSTRAINTS |
|---|---|---|
| **Teachers** | Service delivery agents | Showing up, effective teaching |
| **Students and Families** | Users of the service | Attending school, learning |
| **Schools** | Managers of front-line service delivery | Leading school staff, managing inputs |
| **Governments** | Designers and managers of the system | Designing policy and incentives, allocating resources |

## Methodology

Extensive literature already exists that describes the theoretical underpinnings of RBF.[8] This report will not restate this discussion but rather will present the evidence of when and how RBF can improve educational outcomes. The target audience for this report are stakeholders who are interested in using RBF to unlock binding constraints in the education system to improve learning. A limitation of this report is that we were not able to directly solicit the views of country clients, but we advocate for more analytical work to be undertaken to reflect the perspectives of local actors and stakeholders as they are clearly an important constituency in RBF. There is work underway as part of the REACH trust fund to gather more country-level information about RBF.

For the section on *When Incentives Work*, we reviewed academic research available on the following levels: RBF and teachers; RBF and

7 World Bank (2017)

8 See, for example, Clist (2016), Clist and Verschoor (2014), and Birdsall and Savedoff (2012).

schools; and RBF and students and families. The scope of the review included articles published since 2000.[9] We restricted the search to experimental or quasi-experimental impact evaluations in developing countries, though relevant or seminal research from developed countries is cited where appropriate. The primary instrument to conduct the search was Google Scholar. A number of other studies that had not been identified were added later. These were i) references given in articles found during the initial search, ii) newly published articles that were not available during the initial drafting of the document, and iii) other articles pointed out by colleagues and reviewers. A total of 41 impact evaluations were included. Additionally, 8 reviews of the existing evidence were included in the search.[10]

After the initial search, papers were kept or discarded based on whether the intervention evaluated fell under the three RBF levels. Papers were then classified by theme and coded for similarities and differences, including a note on effect sizes (when possible). A conceptual framework for each topic was outlined, and gaps in the literature were identified. The goal of this literature review was to provide greater context for the findings from the REACH grants, and to underscore how they contribute to the evidence base. Strictly operational evidence plays a smaller role in the *When Incentives Work* section because impact evaluation evidence was available.

Unfortunately, there is less academic evidence available on the impacts of *RBF and Governments*, which drove the decision to focus this section on operational evidence. Nonetheless, quasi-experimental and experimental evaluations were added where available. Finally, both sections draw from the theoretical literature on the relevant topics.

### Methodology to Collect Operational and Tacit Knowledge

The operational information that we use in the report comes from our desk review of documentation of projects in the education sector that have RBF components and were either cited by survey takers as good examples to review or stood out as flagship projects such as Big Results Now in Tanzania. We also conducted a survey of 46 staff from development agencies who design, implement, and evaluate RBF interventions and programs in the education sector to elicit their opinions on RBF in education, keeping their responses anonymous to encourage candor. We also carried out follow-up interviews with 19 of these staff to supplement in more detail the findings of the survey. These interview responses were also anonymous. We developed the survey questions based on the experiences from REACH, which generates evidence and knowledge on RBF in education.

9 Though most of the RBF literature was published in the 2010s, some of the impact evaluations of conditional cash transfers date back to the early 2000s and even late 90s. We chose to only include articles published after 2000 to narrow down the scope.

10 Murnane and Ganimian 2014, Bard et al 2014, Evans and Popova 2016, Glewwe and Muralidharan 2016, McEwan 2015, Molina-Millan et al 2016, Fiszbein and Schady 2009, Results for Development 2016, Snilstveit et al 2015, Imberman 2015.

### What is REACH?

Results in Education for All Children (REACH) is a program housed at the World Bank that supports efforts to improve education, especially for the most vulnerable populations, by helping country systems focus on results. It was established in 2015 and currently funds 33 RBF activities in education in 23 countries around the world. REACH also provides technical support and advice on RBF in education to World Bank teams and other development partners. The main purpose of the program is to contribute to the evidence base around RBF in education.

*Table 2: Sources of Information*

| SOURCE | NUMBER | AGENCIES / TOPICS |
|---|---|---|
| **Impact evaluations (quasi-experimental and experimental evidence)** | Total: 42 papers [11] | RBF with Teachers: 16<br>RBF with Students and Families: 13<br>RBF with Schools: 6 |
| **Meta-analyses** | 8 reviews of evidence and meta-analyses | Several topics: 5<br>RBF with Teachers: 1<br>RBF with Students and Families: 2 |
| **Project documentation** | Total: 20 documents (a mixture of project design documents, status reports, completion reports, and evaluations) | ADB, DFID, GPE, IDB, WBG<br>RBF and Governments: 20 |
| **19-question survey** | 46 respondents | ADB, SIDA, DFID, GPE, WBG, independent evaluation firm |
| **Follow-up interviews** | 15 interviewees | ADB, SIDA, DFID, GPE, IDB, WBG |

### Survey Design

To capture practitioners' perceptions and insights, we created a 19-question survey that aimed to define the attitudes and behavior of staff in development agencies. The survey was sent to roughly 200 individuals who work in international development in the education sector, though not necessarily with any experience with RBF.

The agencies that were chosen to participate in the survey were those that have funded and/or implemented operations in education that have used RBF. Many of these agencies have had experience with RBF through projects for which the World Bank has been the main implementing agency.

11 Some covered several areas.

We created distribution lists for the survey and emphasized that anyone could opt in to take the survey. There were 46 respondents from five development agencies — the Asian Development Bank (ADB), the Swedish International Development Cooperation Agency (SIDA), the UK Department for International Development (DFID), the Global Partnership for Education (GPE), and the World Bank Group (WBG) and one independent consulting company that has done evaluations of RBF programs for the DFID. We subsequently conducted 15 follow-up, semi-structured interviews with staff from six development agencies — the ADB, SIDA, DFID, GPE, the International Development Bank (IDB), and the WBG. To ensure that the survey responses were focused on the education sector and were practitioner-oriented, the respondents were asked whether they had designed, implemented, and/or evaluated RBF activities in education. The respondents could choose from multiple answers: 34 indicated they had designed RBF activities, 28 had implemented them, and 15 had evaluated them. Two respondents indicated that they were starting an RBF project, another two wrote that their work was not related to RBF (so it is unclear why they took the survey), and a final pair wrote that they oversaw RBF activities as part of their portfolio. The survey tool did not allow for matching responses, so it is unclear whether those who said they designed activities also implemented them or evaluated them. However, the data still showed that most of the respondents had expertise in the design and implementation of RBF activities in education.

The survey data are purely qualitative and may be skewed given that respondents self-selected into taking the survey, in other words, those who completed it may have felt more strongly than average about RBF for positive or negative reasons. In addition, the survey only captured the viewpoints of staff of development agencies, not of officials in national-level ministries and administrative units or other recipients and implementers of RBF.

Wherever possible, we have bolstered this qualitative information with supporting evidence from other sectors (the most comparable social sector being health) and with other published studies.

RESULTS-BASED
FINANCING IN EDUCATION

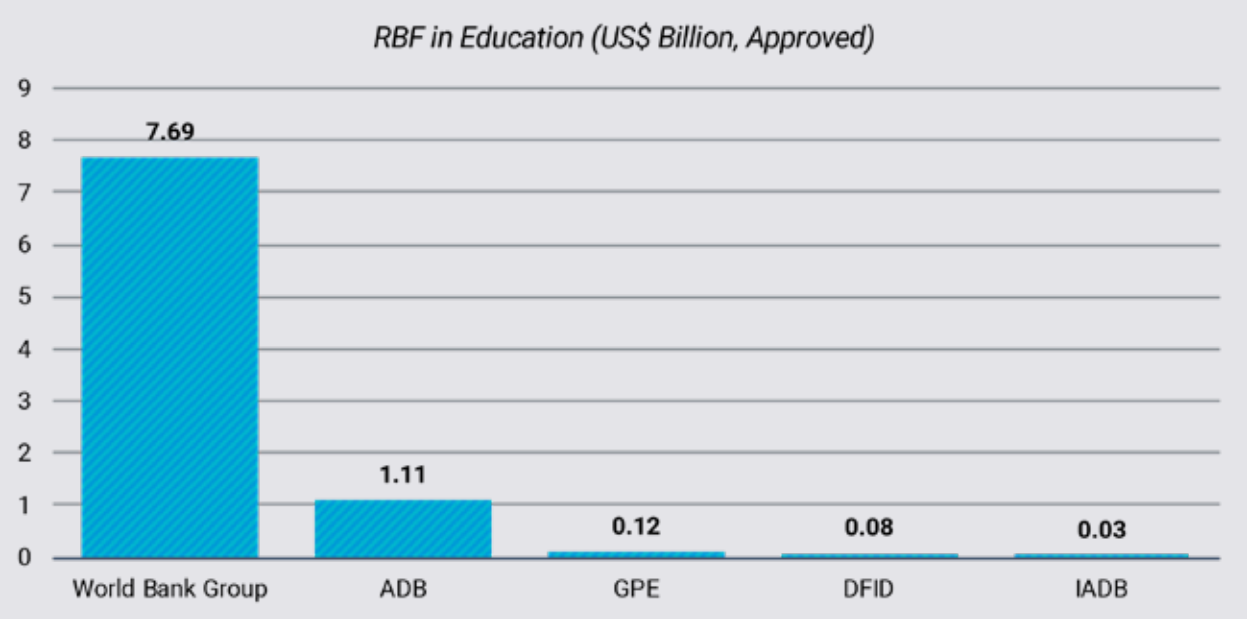
# Donor Portfolios

RBF is a relatively recent phenomenon. Although the first interventions appeared in the 1990s, they were slow to take off. Beyond several pilot initiatives, most of the growth in RBF projects has happened over the past decade. The Bank launched its Program for Results (PforR)[12] instrument in 2012 (though it had been using investment project financing with disbursement-linked indicators even earlier), and the Asian Development Bank (ADB) launched its six-year pilot of an RBF instrument in 2013.

More recently, the African Development Bank has proposed creating a lending instrument based on investment project financing (IPF)[13] and disbursement-linked indicators (DLIs).[14] Nonetheless, the World Bank remains the primary funder of RBF initiatives in education, both as a direct lender and through other implementing agreements with other donors such as the GPE.

Figure 1 presents a summary of the current results-based programs of some of the largest aid donors. Note that the World Bank Group also manages trust funds and other lending mechanisms financed by third party donors. Since these are hard to track down, results are shown by ultimate lender of the funds allocated.

***Figure 1: Education RBF Portfolio of Major Donors (2014–2018)***




12 PforR or Program for Results is one of the three World Bank financing instruments. What sets it apart from other instruments is its focus on results. It uses a country's own institutions and processes and disburses funds against the achievement of a series of agreed-upon results.

13 IPF provides a loan or credit/grant financing to governments for activities that create the physical/social infrastructure necessary to reduce poverty and create sustainable development.

14 Disbursement-linked indicators (DLIs) provide the government with incentives to achieve key program milestones and improve performance.

### ADB (Asian Development Bank)

Since July 2014, the ADB has committed a total of US$4.28 billion in RBF projects. Of these, US$1.11 billion (or 27.9% percent of the total) is committed in the education sector. Only the energy sector has a higher amount of RBF lending from the ADB (with US$1.5 billion). The ADB was one of the pioneers in results-based financing, introducing it in 2013 for a six-year trial period. The most recent evaluation of the pilot (published in November 2017) suggested that the pilot had had generally positive results. Despite some delays because of a lack of familiarity by among the implementers with the design and execution of these projects, RBF projects have been rolled out effectively. Key stakeholders (in both governments and agencies) have endorsed the lending instrument, and as a result, demand for RBF is expected to grow.

### AFDB (African Development Bank)

The African Development Bank approved its results-based financing instrument in November 2017. As yet, there are no projects that are using RBF.

### DFID (Department for International Development)

The UK's DFID has been active in using results-based financing for almost a decade. Since 2011 there has been a strong focus on the value for money component of RBF, to the point where it is often associated with DFID.[15] There have been three main education projects based on RBF. The first was a 2011 pilot in Rwanda that rewarded schools for the number of their students who completed education. The second was a 2012 project in Ethiopia that rewarded the government for the number of students who took and passed a graduation exam. The lion's share of the DFID's RBF budget is spent on the Girls' Education Challenge (GEC), a global RBF initiative to fund projects aimed at increasing girls' access to education. The budget for the GEC alone is around US$454.4 million (not shown in the graph because it was approved before 2014).

### GPE (Global Partnership for Education)

The Global Partnership for Education (GPE) adopted a new results-based funding model in 2014. Up to 70 percent of GPE grants are disbursable following the adoption by the government of an education sector plan and a commitment to increase education spending and to improve data collection and analysis. This part is not strictly contingent on results. However, the remaining 30 percent of the GPE grant is paid only if targets are met. Disbursement is contingent on improvements in three dimensions: equity, efficiency, and learning outcomes. As of February 2018, the total GPE grant portfolio was around USS1.8 billion, of which US$130 million (or 7 percent) was part of the variable tranche and thus results-based.

### IDB (Inter-American Development Bank)

The IDB pioneered a results-based instrument called performance-driven loans in 2003. Several loans were approved, but the results were mixed.[16] As a result, the instrument was discontinued in 2009. Among the reasons for this was the fact that there was little demand because of the strict requirements related to disbursement.[17] Also, the verification of outcomes caused disbursement delays because outcomes had to be matched to specific expenditures. In 2017, a new results-based program (*Programa basado en resultados or* PBR) was piloted. This instrument corrected the

15 ICAI (2018)

16 R4D (2016)

17 IDB (2014)

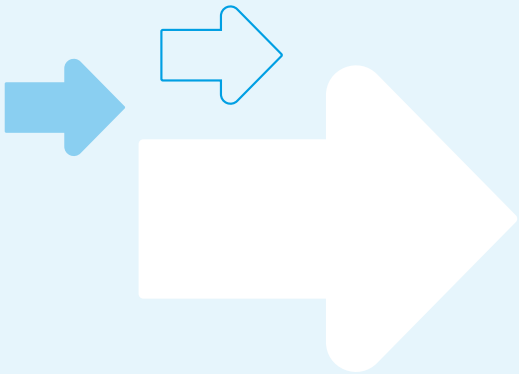
perceived problems of PDL, in that it disburses against outcomes instead of expenditures and allows partial disbursements against partial outcomes among other changes. As of 2018, one education project is being implemented under this instrument for a total of US$30 million.

### NORAD (Norwegian Aid Agency)

The Norwegian Aid Agency (NORAD) has participated in results-based financing initiatives in three key areas: health, climate and forestry, and clean energy.[18] Norway has committed around NOK 2.1 billion to the World Bank-managed Health Results and Innovation Trust Fund (HRITF) and another NOK 1.1 billion to Gavi (the Vaccine Alliance). Norway has also channeled NOK 6.4 billion bilaterally (to Brazil and Guyana) through the Norwegian International Climate and Forest Initiative. However, there are no significant investments in education RBF as of today.

### SIDA (Swedish International Development Cooperation Agency)

The Swedish International Development Cooperation Agency (SIDA) has been active in RBF for a few years in a number of areas. However, it does not publish centralized portfolio data so it is hard to gauge the amount of funds committed to RBF.



18 NORAD (2015)

# Part I – When Do Incentives Work?

**Results-based financing is based on the idea that incentives can help individuals and agencies in the education sector to work towards improving learning outcomes for all.** Results-based financing can take many different forms. Teacher incentives and performance grants focus on service delivery agents (teachers), and organizations (schools), while interventions such as cash transfers focus on the users and recipients of the service (students and families). These topics were selected because they have been the most commonly researched topics when it comes to RBF and education, and because they showcase several key features of RBF at sub-national levels.

**In this section, we explore each type of incentive using the existing, global evidence base** and discuss the key factors to consider when planning an RBF intervention. Key findings from REACH-funded grants will be used as case studies that add to the academic literature and illustrate the practical challenges of designing and implementing an RBF intervention.

The choice of the design issues to discuss is driven by the available evidence. There are many design choices that have unfortunately not been studied yet, and this limits the analysis. Some of these possibilities for further research are mentioned later on.

The following table shows some of the overarching design issues that are relevant to all interventions. As we move through each section, we will highlight those particularly relevant to the level in question (teacher, students and families, and schools).

## Design Issues

*Table 3: Design Issues to Consider*

| | |
|---|---|
| **The incentive scheme** | How will the scheme itself work? What metrics will be chosen as indicators? Does complexity matter? Do conditionalities matter? |
| **Who to incentivize?** | What actor should be incentivized? Sometimes the relevant actor is obvious (such as with teacher incentives), but in many cases there are several options available. |
| **What to incentivize?** | Should one incentivize final outcomes (such as learning) or intermediate outcomes and outputs (such as teacher attendance or student enrollment)?<br>Should one incentivize at the individual level (for example, individual teachers), or at the group level (schools)? |
| **Behavioral responses** | What unforeseen behavioral responses could appear? Is this likely to change the perceived beliefs, preferences or identities of the agents involved?<br>Is this likely to cause gaming and cheating? |
| **Sustainability** | Will the effects of the intervention last?<br>Is the intervention financially sustainable? |
| **Complementarities** | Should one combine RBF interventions with other interventions? Are the effects additive? |

## Using RBF with Teachers

Teacher incentives are schemes that reward teachers for their performance. The rewards are usually cash, but sometimes they can be in-kind (such as a bag of rice) or intangible (for example, a certificate of recognition). Incentive schemes can be designed in many different ways. For example, incentives can be individual or group-based, and they can be linked to the attendance or performance either of the students or of the teachers themselves. The rationale behind these interventions is that a conditional reward will lead to increased teacher effort, which will lead to improved student outcomes.

### The Theory Behind Using RBF with Teachers

The theory behind paying teachers for their performance is based on personnel economics and compensation theory. Under a contract that pays a fixed salary, agents have no incentive to supply effort[19] since compensation is not contingent on an output. However, linking payment to some sort of output or outcome (such as student results or teacher attendance) will theoretically induce teachers to supply more effort and therefore increase or improve that output or outcome.[20]

This improvement in results can happen through several channels. One of them is simply higher teacher attendance, which is

19 In the simplest model, beyond the minimum effort threshold under which they will be fired.

20 For example, see Lazear (2003).

especially relevant in developing countries where teacher absenteeism levels are often quite high.[21] High rates of teacher absenteeism obviously hinder student learning. Therefore, schemes that incentivize teachers' effort and lead to better attendance can lead to better results simply by increasing the number of hours of teaching. Of course, teachers can improve their attendance while the amount of instruction time stays the same (for example, if teachers allocate their time to administrative tasks or are present in schools but not in the classroom). However, there are also other ways for results to be improved. The incentives might induce increased effort from those teachers who already show up by making them more motivated (for example, if teachers felt that they were not being valued for their contributions) or by making teachers fear dismissal. Both of these impulses could induce teachers to spend more time on teaching, to make the content of their teaching more effective, and in general to engage in other strategies to improve student learning.[22]

### Does Using RBF with Teachers Improve Outcomes?

The evidence on teacher incentives as a whole is mixed. A vast review conducted in 2016 found that teacher incentives do not qualify as one of the education interventions that consistently improve student outcomes.[23] Some teacher incentives schemes seem to improve student performance, even substantially,[24] while others have no effect.[25]

In general, the results for developed countries are the most disappointing.[26] A theoretical reason is that salaries in those countries are already relatively high and, therefore, higher incentives would be required to get a significant behavioral response. However, some incentive schemes with modest bonuses have managed to elicit large responses, so the relative size of the incentive may not be the main factor behind these differences between developed and developing countries.[27] It is also worth mentioning that increasing teacher salaries unconditionally does not lead to better student outcomes whatsoever.[28]

The results for interventions in developing countries are somewhat more positive, and the effects are larger. A recent meta-analysis found the effect size of teacher incentives on student learning in developing contexts was around 0.08 SD for math and 0.00 SD for language.[29] In general there is a wide range of results, with some interventions reporting large effects[30] and some reporting smaller or even negligible effects. Nonetheless, the evidence base is still limited by the small number of interventions that have been rigorously evaluated.

The best way to reconcile these divergent findings for now is to recognize that design and context matter a lot. Table 4 lists a series of crucial issues that affect the design of teacher incentive interventions.

21 Duflo et al (2012)

22 Glewwe and Muralidharan (2016)

23 Evans and Popova (2016)

24 See, for example Muralidharan and Sundararaman (2011) in India or Lavy (2009) in Israel.

25 Or even a negative effect; see Fryer (2013) for an intervention in New York City public schools.

26 Imberman (2015)

27 Murnane and Ganimian (2014)

28 De Ree et al (2015)

29 Snilstveit et al (2015)

30 For example, Muralidharan and Sundararaman (2011) reported a 0.28 SD improvement in math and a 0.16 SD in language.

*Table 4: Factors that Can Affect Teacher Incentives*

| FACTOR | EFFECT |
|---|---|
| **Size** | The size of the incentives does not seem to have a large impact on the size of the effect. |
| **What to incentivize** | Incentivizing attendance can improve attendance if monitoring and accountability mechanisms are in place, but has mixed effects on learning outcomes.<br>Incentivizing learning outcomes has mixed effects on learning outcomes. Some interventions have yielded substantive improvements, but others have not. Context and design seem to be key. |
| **Who to incentivize: Individual or group-based** | Both individual and group-based incentives have been shown to have positive effects, though the latter tend to be smaller. |
| **Which metrics to use: level, gains, and percentiles** | Unclear. Some report "pay-per-percentile"[31] is more effective at raising student scores across the distribution than other simpler schemes (like levels or gains), while others report that it has similar results to paying for learning gains or level attained. There may be a trade-off between design complexity and ease of use. |
| **Behavioral responses: loss aversion** | There is some evidence from the US that loss aversion (receiving a bonus upfront and losing it if learning outcomes do not improve) can induce higher levels of effort, but more evidence needed. |
| **Behavioral responses: high stakes and uncertainty aversion** | Evidence from other fields suggests that high stakes incentives (for example, high enough to induce significant volatility in income) can decrease performance and make agents more risk-averse. |
| **Behavioral responses: information** | Evidence from other fields suggests that information in the form of labeling or framing can have an effect on how agents understand an incentive. For instance, framing an experiment as a "Community Game" can make agents more cooperative.[32] |
| **Gaming and cheating** | Incentives can induce agents to cheat by altering the results. This can create the illusion of improvements in learning outcomes that disappears when assessed using a different instrument or test. |
| **Sustainability** | Teacher incentives can be designed to be cost-neutral, but some agents will be net losers and will be likely to oppose the scheme. |
| **Long-term effects** | Besides changing the way current teachers behave, incentives can also change the applicant pool for future teachers. There is limited evidence that introducing incentives could attract more qualified teachers, at least based on their grades. |

*Source:* Authors' summary

31 "Pay-per-percentile" rewards a group of students' ranking vis à vis a comparable group of students, whereas "learning gains" rewards increases in test scores, and "levels" rewards reaching a certain threshold (for example, a passing test score).

32 Gneezy et al (2011)

### Design Issues: What to Incentivize (Outputs or Outcomes?)

A design question that comes up quite frequently is what to incentivize in a teacher incentive scheme. The overall objective is to improve learning outcomes, but as previously mentioned, this can be achieved through different channels and mechanisms. Traditionally, two types of incentives have been used: those that reward effort (for example, teacher attendance) and those that reward outcomes, which were just discussed (for example, student results).[33] In RBF terms, the latter option involves rewarding results that are further along the results chain (an outcome), whereas the former would be rewarding an intermediate outcome.

In the case of teacher incentives, some evidence suggests that rewarding attendance can increase teacher attendance. Teacher absenteeism can be very common in many countries, and this directly undermines student learning. In an intervention in Rajasthan, India, researchers found that paying teachers for their daily attendance reduced absenteeism by 21 percentage points and increased test scores (0.17 SD).[34] To make sure teachers were actually present in the classroom, schools were given cameras with tamper-free timestamps. Every day, students were instructed to take a picture of the teacher both at the beginning and at the end of the school day.

However, attendance incentives do not seem to have a significant effect on student learning outcomes (and occasionally even on attendance itself) as several other interventions have failed to show any positive effects. A crucial factor seems to be the accountability mechanism. One of the successful programs that has been evaluated[35] used cameras, whereas other programs delegated accountability to either the school (for example, the principal) or the community. This can be an attractive alternative because cameras are expensive and can still be tampered with. However, delegating accountability to schools can be ineffective since there is a risk of collusion between teachers and other stakeholders. An intervention in Kenya gave principals bonuses if they reported teacher absenteeism,[36] but the program had no effect on attendance or student learning (in fact, the principals reported enough missing teachers to get the bonus but no more).

A different program in Uganda tested a scheme to crowdsource attendance reporting from principals and parents and concluded that giving bonuses to principals but not to parents led to somewhat improved teacher attendance and higher reporting of absent teachers.[37] However, both principals and parents systematically under-reported absences, which once again suggests that delegating attendance reporting to schools may be cheaper but somewhat ineffective.

### Design Issues: Who to incentivize?

***Individual or Group-based***

Another design question is whether incentives should be individual or group-based. In general, the evidence seems to show that both individual and group incentives can work, though the latter tend to have smaller effects.[38] In other sectors, group incentives have also been shown to work.[39] Theoretically, individual incentives should work better because they connect a person's effort with a reward whereas group effort is beyond the individual's sole

33 Murnane and Ganimian (2014)

34 Duflo et al (2012)

35 Duflo et al (2012)

36 Cited in Murnane and Ganimian (2014)

37 Cilliers et al (2013)

38 Murnane & Ganimian (2014), Glewwe & Muralidharan (2016), Lavy (2002), Lavy (2009)

39 Garbers and Konradt (2014)

control though arguably individual student learning is also beyond the teacher's control. However, there is some evidence that group incentives may still work. For example, a study in the Indian state of Andhra Pradesh found that rewarding teachers for group-level improvements led to improvements in student learning similar to those seen for individual incentives. However, after two years of implementation, the individual group had shown greater improvements.[40]

### Design Issues: Which Metrics to Use

***What Metric to Use: Level, Piece Rate, or Rank***

Finally, in order to incentivize student outcomes, the question arises of what specific metric to choose. There are broadly three options: by levels, piece rate, or rank. An incentive scheme based on levels will reward teachers by the number of students hitting a certain target (or level). For example, an intervention can provide a bonus to teachers based on the fraction of students who pass an exam. While these systems are easy to understand, they can also provide some perverse incentives.[41] For example, if the system rewards teachers for the number of students who pass a test, then they are given no incentive to help students at the bottom end of the distribution (because it is unlikely that they will pass). Instead, teachers may feel that it makes more sense for them to focus on students around the middle of the distribution. However, this can be mitigated by including several different thresholds in the design of the incentive. Thus, piece-rate systems instead reward teachers for incremental improvements instead of for reaching a single threshold (for example, the total increase in the test scores of students in a class). Finally, rank-based incentives provide bonuses based on the teachers' ranking vis à vis the rest of the universe (for example, the other teachers in the district).[42] A key aspect of a rank-based system is that it is cost-predictable because one teacher's percentile gain is another teacher's percentile loss. This may be attractive to administrators because it makes it easier to gauge the fiscal impact of a program.

A recent study proposed a hybrid method named pay for percentile.[43] This method rewards teachers for their students' ranking position in comparison with an equivalent group of students defined in advance. Therefore, it combines the features of a piece-rate system and of rank-based incentives. It is similar to piece-rate systems in that an improvement is "worth" the same at every point of the distribution and is similar to rank-based systems in that improvements are measured by percentiles (relative position) instead of scores.

An evaluation of an intervention in China found that a pay-for-percentile scheme did indeed work better than two similar schemes that rewarded teachers based on a class average of the level reached by students or by their gains, as measured by a test.[44] Despite the greater complexity of the intervention, teachers understood it and reacted accordingly. While in the levels and gain treatment groups, the teachers mostly focused on the students whom they thought could improve the most, in the pay-for-percentile scheme, they increased the coverage and intensity of their instruction for the class as a whole.

To look at whether these effects have been replicated elsewhere, REACH funded an evaluation in Tanzania that involved two interventions, one that used learning gains and one that used pay for percentile.[45] The first intervention rewarded teachers based on how many students attained a certain level or score in a test. However, to avoid the perverse

40 Muralidharan & Sundararaman (2011) and Muralidharan (2012)

41 Murnane and Ganimian (2014)

42 Imberman (2015)

43 Barlevy and Neal (2012)

44 Loyalka et al (2016)

45 Mbiti et al (2018b)

incentives just discussed, the intervention created different levels across the distribution. This partly solved the problems of simple models with just one or two levels (where teachers would have no incentive to help the students that fall far away from the threshold).

Contrary to the intervention in China,[46] the evaluation of these interventions found that the levels system worked just as well as the pay-for-percentile system. Therefore, further research is needed to tease out whether pay-for-percentile really is a better way of eliciting improvements in teacher performance and student outcomes. A key issue is whether there is a complexity-efficiency trade-off.

It is also worth noting that changes in test scores can be prone to volatility caused by student cohort characteristics or one-time shocks.[47] This means that incentive schemes can be based on noisy metrics that do not accurately reflect reality (and effort). This is especially problematic for smaller schools, since the chances of random variation are much higher in those schools because of their small number of students. Test scores themselves, rather than *changes* in test scores, are less noisy.[48]

### Design Issues: Behavioral Responses

Another critical design issue is the behavioral response of the agents involved. For example, an intervention in Chicago public schools included two different treatments: a teacher bonus based on student performance in an exam at the end of the year compared to a lump sum of money given to teachers in advance that would be taken away if student outcomes did not improve.[49] Interestingly, the latter treatment had a positive effect on student learning, but the former did not. The authors of the evaluation of this intervention suggested that this result may have been due to loss aversion, the human tendency to value losses more than the equivalent gain.[50]

Information can also alter a person's behavioral response. Some evidence suggests that the size of the incentive on its own is not crucial, but that it can interact positively or negatively with other design features.[51] Teacher incentives, both monetary and non-monetary, convey a lot of information beyond the incentive scheme itself. This information can have a very important impact on the behavioral response of teachers, an issue that has been researched extensively in the behavioral economics literature.[52] For example, the introduction of an incentive scheme can make the teachers think that they are not trusted by whoever is in charge of the incentives (perhaps the Ministry of Education), leading to a drop in their morale.

Thus, information can have an effect through what it reveals about the *designer/implementer* and what it reveals about the *situation* itself. An example of the former is when teachers lack motivation because they believe that being offered an economic incentive for effort means that the administrators do not trust them. An example of the latter is when teachers teach to the test because they see the offer of incentives treats teaching as a market exchange rather than a vocation. There are endless examples of cases where the interpretation of a specific design can lead to very different outcomes, which highlights the need to think about the realities of the implementation of a specific program beyond its theoretical design. It is not just the design of an incentive that matters, but the way this is conveyed to the agents themselves. This shows that clearly communicating to relevant stakeholders how an incentive scheme will work and the ideas behind it is incredibly important.

46 Loyalka et al (2016)
47 Barrera-Osorio and Ganimian (2016)
48 Murnane & Ganimian (2014)
49 Fryer et al (2012)
50 Fryer et al (2012)
51 Loyalka et al (2016)
52 See Bowles and Polanía Reyes (2012)

### Design Issues: Sustainability

An additional factor to keep in mind is the sustainability of teacher incentives. The financial sustainability of the incentive program largely depends on the nature of the program itself. It is possible to design interventions that are relatively cost-neutral, or at least cost-predictable.[53] For example, if part of teachers' annual salary increases is instead redesigned as variable pay, the only additional cost will be the administration of the program (for example, the grading of any test used to evaluate students' learning outcomes). However, the political economy of such an intervention is likely to be much more problematic as it inevitably involves winners and losers. Instead, a program that adds a variable pay component to any planned annual salary increases will be more popular but also significantly more expensive.

Chile stands out as an example of how countries have addressed these challenges. Instead of rolling out a teacher incentive system outright, it created a voluntary scheme first and engaged in extensive consultations and negotiations with teacher unions. This gave teachers a chance to get to know the system and choose whether or not to adjust their classroom practices. The government also implemented a series of reforms (such as steady salary increases) that reassured teachers that any losers from the new system would be compensated.[54]

### Design issues: Long term effects

Another question is what will be the long-term effects elicited by the program. There are two channels through which teacher incentives may change teacher behavior: (i) changing the behavior of current teachers, which will happen in the short-term, and (ii) changing the population of people who apply (and then become) teachers, which is likely to become more important in the long-term.

Unfortunately there isn't much evidence on the effect of teacher incentives on recruitment. But theoretically they could alter the pool of applicants, perhaps by attracting more motivated candidates (which would do better under a pay-for-performance scheme) and discouraging less motivated candidates. There is experimental evidence that shows that the way a position is advertised can have drastic effects on the people who are recruited. In Zambia, an intervention tested the effect of two advertisements for the same community worker job. The first ad highlighted career progression as a key component of the job, whereas the second highlighted community service. The career progression ad attracted candidates that did not seem different based on observable characteristics but then went on to perform far better than the other group.[55] Additionally, evidence from the US suggests that teacher incentives could actually attract more qualified candidates (at least judging from grades and test scores).[56] REACH is currently funding research to examine how pay-for-performance teacher contracts in Rwanda affect selection and recruitment.

### Design Issues: Gaming and Cheating

When money (or some other good) is tied to a performance indicator, there will inevitably be perverse incentives for the agent involved. The result is cheating and gaming. For example, teachers could tell their students what to answer when they take high-stakes exams since the results will be used to evaluate the performance of the teacher. Teaching to the test is another (much less flagrant) way of gaming the system. In the developing world,

53 Muralidharan and Sundararaman (2011)

54 World Bank (2017c)

55 Ashraf et al (2014)

56 World Bank (2018e)

there are a few examples of interventions that have led to these kinds of behavior.[57] For example, an evaluation in Kenya showed that incentives improved student scores according to the tests that would be used to determine the rewards. However, when they were tested with a different test that measured the same content, there was no improvement. This suggests teachers may focus on teaching to the test (or drilling) rather than on activities that improve students' content knowledge and skills.[58] Alternatively, teachers may reallocate their time away from subjects that are not linked to incentives (such as art) to subjects that are.

It is difficult but possible to prevent gaming when designing teacher incentives. Generally, it can be done by including several verifiable metrics that aim to reduce perverse behavior. However, the cost involved in adding more indicators is ending up with a scheme that is over-designed and too hard to understand that confuses teachers about what they should prioritize. There is some evidence that when several variables are incentivized, agents choose to invest the most effort in achieving those that are easiest to attain. For example, an intervention in Mexico created several incentive schemes that included both participation in final exams and student outcomes as metrics (to prevent teachers and principals removing low-performing kids from taking the test).[59] This led to increases in participation (the easier metric to pursue) but no improvements in student test scores.

It is also worth noting that, while ideally gaming and cheating should be minimized, they do not necessarily invalidate an intervention.[60] It is possible both for some agents to cheat and for there to be a general improvement in the targeted outcome. In fact, reducing cheating to a minimum comes at a cost and is not always desirable since supervision and monitoring costs money too. Therefore, there are trade-offs involved between the cost of cheating and the cost of supervision, some of which will be discussed later in the report.

### Conclusion

One of the key objectives of using RBF is to focus on outcomes down the results chain, such as intermediate and final outcomes, rather than on inputs as has often been the case with traditional financing. However, the available evidence thus far on teacher incentives is mixed. What has emerged is that interventions that use teacher incentives can be successful given the right conditions, but they can also have negligible or negative effects.

When designing an intervention, it is important to keep in mind the issues discussed in this section such as the structure (and complexity) of the incentive scheme, the behavioral response of all the agents involved, and the possibility of gaming and cheating. Perfecting the incentives at just one level will be of little use if the incentives at the other levels are not aligned too. Therefore, an incentive scheme that aligns the incentives of all agents involved in the education sector will probably work better than a scheme that just influences teachers.

Many gaps remain in the literature. The most relevant are how different incentive structures (pay-for-percentile or levels, for example) operate in different contexts and the fiscal impact and long-term effects (both for students and for teachers) of teacher performance incentives.

57 Kremer et al (2010)

58 Kremer et al (2010)

59 Ganimian and Murnane (2014)

60 Murnane and Ganimian (2014)

## Using RBF with Students and Families

In RBF, incentives usually involve rewarding an individual in exchange for a certain behavior or action. The most popular example of this sort of mechanism are conditional cash transfers, which have been implemented for over 25 years, often quite successfully. In this section, we discuss how to use RBF with students and their families to encourage improved learning outcomes.

### The Theory behind Using RBF with Students and Families

One of the main reasons for providing incentives to students and their families is that households have been shown to underinvest in education.[61] The literature points to several factors that can lead to this underinvestment. The first is the fact that households and students may want to invest in education but may not have the money to do so. In these cases, a subsidy would relax that constraint.[62] A second reason is that households may be unwilling to invest in education because they are not aware of the returns to education. In this case, directly providing a payment in return for some action (such as school attendance) would signal that schooling is important and worthwhile, thus increasing households' awareness of the returns to education. An alternative would be to conduct an information campaign to directly inform families about these benefits. One of the REACH grants, which we will discuss later on, targets precisely this channel. A final reason for providing incentives to students and their families has to do with behavioral factors such as present bias and discounting. Sometimes households would like to invest more in education but end up allocating resources to meet more urgent needs. In this situation, mechanisms that could prevent this, such as conditional incentives, can play an important role.

### Does Using RBF with Students and Families Improve Outcomes?

The evidence base regarding the provision of incentives to students and their families is quite large, though mainly focused on conditional cash transfers (CCTs). The first CCT was PROGRESA, a Mexican program launched in the late 1990s that distributed transfers to families in exchange for ensuring that their children attended school and for taking them to clinics for preventative health checks. The program was also randomized, which has made it relatively easy to evaluate since its inception. Research has shown that, for children aged between 6 and 14 years old in 2003, PROGRESA raised the number of years of schooling by 0.25 (about three months) for boys and 0.32 for girls (about four months) after

61 Underinvesting, for example, in the sense that households have children who end up getting far fewer years of schooling than what would be optimal based on returns to education in the country in question.

62 Glewwe and Muralidharan (2016)

5.5 years of exposure to the program.[63] Since then, the number of CCT programs around the world has increased tremendously.[64] In education, these programs generally provide cash transfers to family members (which member is the final recipient varies depending on the program's design) in exchange for some behavioral change by the household, usually enrolling their children in school or increasing their rate of school attendance.[65]

Overall, the literature suggests that CCTs decrease school dropouts and increase school attendance and completion for children.[66] A recent meta-analysis found that their effect on attendance has been 0.13 SD (-0.12 SD for dropouts) and 0.12 for completion. Furthermore, evaluations of programs in Brazil, Honduras, Malawi, Colombia, and several other countries have all suggested that they have had a positive impact in variables such as re-enrollment, transition to the next education level, labor outcomes, and even health status.[67]

For example, in Malawi, a conditional transfer increased the number of terms during which girls were enrolled in school during a two-year period by 0.535.[68] Other studies have also found that CCTs have continued to increase enrollment in the year after the cash transfer was received. A cash transfer in Colombia has increased attendance by 2.9 to 3.2 percentage points, while re-enrollment the following year increased by 1.1 to 5 percentage points. This program also includes a treatment option where the household receives part of the transfer the year after the child has re-enrolled, and this has yielded better results than the standard CCT, but both are positive and significant.[69] For Brazil's Bolsa Família, some studies have found that the transfers decreased dropout by 3 percentage points and increased enrollment by the same amount.[70] In Honduras, the story is the same where PRAF, a conditional cash transfer, increased enrollment rates by 8 percentage points and decreased the probability of child labor by 3 percentage points.[71]

One of the benefits of having a wide range of interventions to examine is that there is some evidence of the different factors that can affect the impact of a CCT. Table 5 shows several.[72] Some of these will be discussed in more detail further below.

63 Behrman et al (2009) cited in Glewwe and Muralidharan (2016)

64 For a review, see Fiszbein and Schady (2009).

65 Glewwe and Muralidharan (2016)

66 Glewwe and Muralidharan (2016) and Snilstveit et al (2015)

67 For further information, see Murnane and Ganimian (2014) or Glewwe and Muralidharan (2016). For the impact on health status, a good example is Gertler (2004). For Latin American CCTs, Molina-Millan (2016) is a recent survey.

68 Baird et al (2011) cited in Glewwe and Muralidharan (2016)

69 Barrera-Osorio et al (2011)

70 Glewwe and Kassouf 2012 cited in Glewwe and Muralidharan (2016)

71 Galiani and McEwan (2013)

72 Most of these were identified by Murnane and Ganimian (2014)

*Table 5: Factors that Affect Conditional Cash Transfers*

| FACTOR | EFFECT |
|---|---|
| **Conditionality** | Both conditional and unconditional transfers have effects of similar magnitude in general, but some unconditional interventions have smaller effects. |
| **Information and labeling** | Information treatments (for example, providing an attendance report card with the transfer) can have positive effects that complement the transfer itself.<br><br>Some evidence suggests that labeling cash transfers as school transfers can also improve outcomes, perhaps by increasing the salience or perceived importance of education.<br><br>Some information treatments on their own can also have a positive effect, but this is usually smaller. |
| **What to incentivize** | Conditioning transfers on attendance raises attendance, retention and graduation.<br><br>There is some positive evidence on rewarding students to improve learning outcomes (see cell below). |
| **Who to incentivize** | There seem to be no large differences between fathers and mothers (with some exceptions). Giving kids part of the transfer may increase the magnitude of the effect.<br><br>Some evidence suggests that rewarding students for effort (input) and grades (outcome) can improve learning outcomes, though more evidence is needed. Rewarding goal-setting does not seem to improve learning outcomes. |
| **Other factors** (from Murnane & Ganimian 2014) | |
| **Share of students enrolled** | The lower the initial share of students enrolled, the higher the effect on enrollment. |
| **Size of transfer** | Larger transfers do not always cause larger effects; diminishing returns. |
| **Timing of transfer** | Delaying part of the payment and making it conditional on next grade enrollment increases retention.<br><br>There is limited evidence that more frequent student rewards improve learning outcomes more than one large reward (see "Who to incentivize" section). |
| **Age and grade of recipient** | CCTs more effective in transitions from primary to secondary and from lower secondary to higher secondary. |
| **Poverty level** | The poorer the beneficiaries, the larger the impact. |

*Source:* Data from Murnane & Ganimian (2014) with some modifications and additions from the papers cited in the report

The evidence in terms of attendance is quite clear, but it seems that these incentives do not generally improve student learning outcomes. The overall effect in a recent meta-analysis was measured as 0.01 SD for a composite language/math score, which is indistinguishable from zero.[73] Of course, there are some exceptions. In the case of Malawi, test scores improved under the cash transfer. English test scores improved by 0.13 SD, and math scores by 0.16 SD.[74] In Nicaragua, there were significant gains in the math (0.17 SD) and language (0.23 SD) test scores for young men exposed to the program.[75] However, most other recent evaluations have shown no significant changes. For example, an evaluation funded by REACH in Mozambique found that the gains in attendance in all treatment groups did not translate into improved student learning. This makes sense intuitively because CCTs increase the attendance of the most vulnerable students, who generally have lower grades. Therefore, by virtue of the changes in composition of the student body, it could be expected that the average grade would decrease.

### Design Issues: The Role of Conditionality

There are other questions that remain to be answered, mostly related to the behavioral mechanisms by which CCTs operate. Some evidence suggests that conditionality is not necessary, for instance. An evaluation of an intervention in Burkina Faso that included both conditional and unconditional cash transfers found that both led to similar increases in enrollment.[76] However, the authors also noted that unconditional cash transfers were worse at increasing the enrollment of what they called "marginal children," In other words, children who were not prioritized by their parents in terms of school attendance, such as girls or younger siblings.

In Malawi, however, the unconditional transfer led to only 43 percent of the reduction in dropouts as the conditional transfer, which suggests there are more issues involved than just financial constraints.[77] This was also seen in other interventions.

The overall verdict so far is that *both* conditional and unconditional cash transfers can have important positive effects on attendance. However, it seems that conditional transfers have a greater effect.[78] Additionally, depending on the context, the effects of an unconditional cash transfer can vary. As mentioned previously, in Burkina Faso, marginal children were better off with a transfer that required parents to take them to school. Therefore, it is possible that subgroups are affected in different ways by conditional and unconditional transfers, though as of now there is not enough evidence to tell how or why.

### Design Issues: The Role of Information and Labeling

As mentioned above in the teacher incentives section, an intervention conveys much more than money or goods. It also provides information to the recipients. For example, it may signal what is considered important by the public or it may show that the returns to education are higher than the family initially thought. These changes in beliefs can alter a family's behavior and their subsequent investment in education.

Indeed, several interventions have shown that just providing information on its own can have a positive effect on behavior. In the REACH grant in Mozambique that was mentioned above, one of the main objectives was to evaluate the impact of information. The authors of the evaluation found that the information content of a conditional transfer

73 Snilstveit et al (2015)
74 Baird et al (2011)
75 Barham et al (2013)
76 Akresh et al (2013)
77 Baird et al (2011)
78 Baird et al (2014)

can have a substantial effect on school attendance independently of the transfer itself. In the authors' experiment, the estimated effect of the information treatment (report cards only) on attendance was as large as 54 percent of the child incentive effect and 75 percent of the effect of the parent incentive, which was impressive given that it cost a fraction of the transfer.

Finally, labeling a transfer can also have an effect. An intervention in Morocco tested the effects of an unconditional transfer labeled as a "school transfer" compared to a similar unlabeled conditional cash transfer.[79] The labeled unconditional transfer decreased dropout by 76%, and the effects were similar for the conditional transfer. When looking at re-entry rates the following year, the labeled transfer actually performed better than the CCT.

This suggests that providing information can have significant effects on outcomes, both when attached to transfers and when provided alone. Therefore, in situations where governments or agencies have financial constraints, providing information could be a strong next best option for increasing school attendance and perhaps improving other indicators.

### Design Issues: What to Incentivize

In principle, there is no reason why cash transfers cannot incentivize other kinds of behavior. Evidence has shown that increasing school enrollment does not always translate into improved learning, so perhaps it makes sense to focus on other kinds of behavior that could plausibly improve student outcomes.[80] One of these kinds of behavior is goal setting. In Zanzibar (Tanzania), REACH helped to fund a field experiment among secondary students in Zanzibar that tested the impact of student goal-setting on their performance and whether this impact differed when reinforced with extrinsic incentives (in other words, non-financial recognition awards for meeting self-set goals). It was found that goal-setting increased students' effort, especially for those who exhibited low to medium ability at the baseline and for those to aspired to higher education.[81] However, this increased effort did not translate into improved student performance within the short time period of the study (eight months). The team also found that extrinsic incentives did not enhance the effectiveness of goal-setting for students.

Finally, there is the possibility of incentivizing results further down the results chain, such as learning outcomes themselves. There is a growing literature that looks at whether paying students for improvements in test scores can work, and the evidence from developing contexts is promising. For example, an evaluation of an intervention in Nepal found that rewarding eighth-grade students for their average performance increased test scores by 0.09 SD.[82] A different intervention in Benin tried out three different incentive schemes: the first paid individual students for reaching a specific performance level on tests, whereas the other schemes incentivized groups of four students to perform better on tests. All incentive schemes led to improvements in test scores, ranging from 0.27 to 0.34 SD.[83]

Other evaluations have sought to compare input and output incentives as a way to tease out what specific factors cause improvements in student learning. An intervention in India looked at two different incentive schemes operating within a math computer-assisted learning platform. The first incentive scheme (the input scheme) rewarded students for every

79 Benhassine et al (2015)

80 Pritchett (2001)

81 Islam, A., S. Kwon, E. Masoon, N. Prakash, & S. Sabarwal (2017)

82 Murnane and Ganimian (2014)

83 Blimpo (2014)

learning module that they completed (including quizzes), while the second scheme rewarded students for their scores in a test administered at the end of the scheme (the output scheme). The students were rewarded with points that they could use to purchase real goods in a virtual store. The output incentive scheme led to a 0.27 SD increase in test scores compared to the control group, whereas the input scheme caused a whopping 0.54 SD improvement.

It is tempting to argue that this input incentive worked better because it rewarded key factors of the learning production function rather than leaving it up to students to decide how best to improve their outcomes. However, that conclusion would be premature. The input scheme rewarded students periodically as they mastered each subject, which may have raised the profile of the incentive scheme and reduced present bias. Since the reward offered by the outcome incentive was concentrated at the end of the intervention, students may have discounted the reward more heavily. If this was the case, then the difference in favor of the input scheme would be due to behavioral responses rather than to the distinction between outcomes and inputs.

### Design Issues: Who Should Be Incentivized?

There is also a question of who should receive the incentive. In the case of CCTs for school attendance, this is almost always the student's family, but within the family, each member has different views, beliefs, and preferences. The differences between spouses is particularly relevant. There is strong evidence that gender plays a role in a range of issues related to financial decisions. Women are more likely to pay back microfinance loans and to invest household resources in health and education.[84] However, when targeting parents for CCTs, it is unclear whether gender has a strong effect. An intervention in Morocco we previously mentioned showed similar results for kids regardless of whether the recipient was their mother or their father.[85] The authors suggest this may be because father immediately appropriated the transfer, and indeed the data shows that a majority of mother recipients were accompanied when they went to pick up the money, whereas most fathers picked it up alone.

What emerges from the literature is that outcomes depend crucially on the rules that govern intra-household decision-making. And this is prone to variation by context. Giving transfers to mothers may make a difference in situations where they have bargaining power in the household, but not otherwise. An alternate view is that giving mothers money may increase their bargaining power vis à vis their husbands.

Recently, it has become clear that children play a role in the decision-making process too.[86] After all, they have preferences that are fairly distinct from their parents. A few recent interventions have explicitly targeted children as the recipients of transfers, sometimes with interesting results. Some studies have found that giving money to parents and toys to children had a similar effect in terms of improving student learning.[87]

The Mozambique grant added to the evidence by directly comparing the provision of similar incentives to parents and children in terms of their respective effects on attendance. All female students in senior primary grades had attendance report cards that were given to their parents at the end of each week. In the first group, if the girl's attendance was at 90 percent or more, she received vouchers that could be used to purchase certain school materials. For the girls in the second group, the monetary amount of the voucher was given to the parents, with the option of purchasing the same school materials that were made available to the first

84 Ashraf (2009)

85 Benhassine et al (2015)

86 Dauphin et al (2011)

87 Berry 2015.

group. In the third group, girls simply received the attendance report card with no incentives attached. There was also a fourth control group. The evaluation of the program found that the incentive given to the children was at least as effective as the incentive given to their parents. In fact, the effect was 38 percent higher for the children's incentive group, but the intervention lacked enough statistical power to establish significance.

In conclusion, the effects of an incentive can be different depending on who receives it. Women recipients (often mothers or grandmothers) have been shown to make higher investments in family, health, and education than male recipients. Providing transfers to the students themselves is not yet a widespread practice, but the examples so far have been encouraging.

### Conclusion

There is much available evidence on the effects of CCTs, and overall the results are positive. Giving incentives to students and their families in the form of transfers can increase intermediate outcomes like school attendance and enrollment rates but can also increase final outcomes like graduation rates. The effects of CCTs on student learning are less impressive. While some interventions have shown promise, generally CCTs have not been found to have a positive effect on learning as measured by test scores. It remains unclear whether transfers need to be conditional. The effects of providing conditional and unconditional transfers are similar, though they tend to be larger for conditional programs.

There are still many research gaps to fill regarding context-specific effects, such as the role played by social norms in the households' response to CCTs. For example, in Burkina Faso, an evaluation found that unconditional transfers led to the exclusion of marginal children, whereas conditional cash transfers did not. One possible explanation for this is that the conditional transfer explicitly broke the norm that marginal children were not expected to go to school. The role played by the use of information in interventions (such as the returns on education or the simple labeling of a transfer) also needs further elucidation. Occasionally, it seems that labeling a transfer as an "education transfer" is enough to promote attendance even without any conditionality.

## Using RBF with Schools

RBF with schools usually takes the form of school grants. For the sake of the report, we take grants to be public funds transferred to schools to cover operational (and other) expenses, over which schools have some discretion. This discretion over the allocation of resources is a key feature that sets grants apart from regular school financing (in the form of earmarked transfers to pay for teacher salaries, for example). There are several types of school grants, including those that do not have strings attached (unconditional) and those that do (conditional). Conditional grants include performance-based grants, which are a type of RBF. We will be focusing on these throughout this section. It belongs to the kind of financing policy that incentivizes the front-line of the education results chain (usually schools, which are the direct providers of education services).

### The Theory behind Using RBF with Schools

The main idea behind school grants is that many schools both know how to and would like to improve student learning but often lack the resources or motivation necessary to do so. For those who lack the resources, an increase in financing through grants could help them to implement the improvement plans that they deem appropriate and that would eventually improve learning outcomes. The argument in favor of providing these grants is that school leaders have more knowledge about the deficiencies of their school than planners and officials in any line ministry, so they will spend the money more effectively. However, for those school

leaders who lack motivation, a conditional grant program could induce them to improve their management practices by offering their schools more resources contingent on the school's performance. Indeed, research has shown that school management practices vary widely and that good management practices are associated with better learning outcomes.[88] Since in principle, good management practices can be adopted by lower-performing schools, this could plausibly lead to improved student learning.

### Does Using RBF with Schools Improve Outcomes?

There is only limited evidence on the effects of school grants on improving learning outcomes. One reason for this is that they are rarely stand-alone policies. Many grant programs are the result of the abolition of school fees when schools are compensated for this lost revenue.[89] Other grant programs are bundled together with wider school-based management interventions, which often include training for principals and other staff, or the creation of school committees. This makes it hard to disentangle the effect of each component.

According to the available evidence, the overall effect of grants on learning outcomes is mixed.[90] A recent meta-analysis found that the pooled effect of these grants on a composite language/math score was -0.01 (and not statistically significant), with a range from -0.34 to 0.15. For other outcomes, such as enrollment and participation rates, the results are somewhat more positive. For instance, a recent review found an effect of 0.05 SD in school completion, -0.02 SD on dropout zero, though both results were not statistically significant.[91] Effects on enrolment and teacher attendance were negligible.

Nonetheless, there is plenty of variation amongst the results.Finally, one evaluation looked at the interesting question of whether it makes a difference when schools receive unannounced grants as opposed to grants that are expected.[92] Using data from Andhra Pradesh in India and from Zambia, the authors found that unannounced grants led to improvements in student learning but announced grants did not. In the case of India, this amounted to improvements of 0.08 and 0.09 SD in language and mathematics test scores respectively (for a grant of US$3 per student). In Zambia, the government assigned block grants that also cost a little under US$3 per student. The evaluation found that test scores in language and mathematics both increased by 0.10 SD. The authors suggest that the reason for this was that households offset the anticipated grants by reducing their own spending on education.

For example, a one-time school grant combined with an intervention to improve school management in the Gambia led to a 21 percent drop in student absenteeism and a 23 percent drop in teacher absenteeism but no improvement in learning outcomes.[93] However, the group of schools that only received the one-time grant saw no improvements in any category. A grant and training program for parent associations in Mexico led to reductions in grade repetition and grade failure of 4 to 5 percentage points.[94]

88 See for example Bloom et al 2015.

89 Al-Samarrai et al (2017)

90 Note that this is based on the pooling of school-based management interventions. An overwhelming majority of these include school grants, but they often also include capacity building.

91 Snilstveit et al (2015)

92 Das et al (2013)

93 Blimpo et al (2015)

94 Gertler et al (2012)

*Table 6: Factors that Can Affect Performance-Based School Grants*

| FACTOR | EFFECT |
|---|---|
| **Competitive distribution of resources** | It is unclear whether using a competitive allocation system for school grants makes a difference, since research is lacking. An intervention in Senegal led to improved student outcomes, but results in Indonesia were mixed. |
| **Equity issues** | Equity considerations will arise when including a competitive system for allocating resources. High-performing schools may be more likely to receive the grant, thus increasing inequality. This can be mitigated by creating different competitions based on the socioeconomic backgrounds of the district, for example. |
| **Household response** | There is some evidence that households may reduce their own educational spending if they anticipate an increase in school grants, which could limit the effectiveness of the intervention. Some alternatives could include financing inputs that are harder to substitute for, or providing larger grants. |
| **Long-term effects** | No evidence on this so far, but other interventions in health suggest that short-term grants can help permanently overcome organizational constraints (and thus improve outcomes). |

*Source:* Authors' summary

### Design Issues: Competitive Distribution of Resources

A key design issue is whether to allocate resources competitively or not. For instance, some grant programs may distribute extra funding to schools that meet a series of requirements, such as improvements in student learning or the submission of an improvement plan. The evidence so far is mixed on whether this is effective, though the number of studies is very limited.

In Senegal, a competitive grant program had positive effects on student learning, especially for schools that spent the funds on human resources rather than school materials.[95] Schools could apply for funding for specific projects of their choosing, and the amount of the grant was a sizable US$3,190, or around 7 percent of each school's total annual budget (including teacher salaries). The Ministry of Education's guidelines specified that schools' grant applications had to be focused on pedagogical improvements and to be prepared by a committee of local officials, parents, and teachers. (This is reminiscent of what was found in Indonesia (see below) where increasing links between schools and local officials led to improved student learning.) The evaluation of the Senegal program found that the grants increased student test scores by 0.09 SD after two years.

The REACH intervention in Indonesia used an RBF-based reform of the entire system to try to evaluate the effects of a new performance-

95 Carneiro et al (2016)

based school grant. The intervention created a bonus grant for the top 25 percent best-performing schools in the system, with schools competing against other schools in their district, which reduced equity concerns since their socioeconomic backgrounds were similar.[96] The bonus grant was equal to 20 percent of the fixed grant, which is a sizable amount. The fixed grant was US$4.5 per student for primary schools and US$8.2 for junior secondary schools.

The results, however, were mixed. The test scores of junior schools improved, but those of primary schools fell (though this effect was temporary). These changes in test scores occurred before the new system itself was implemented, which suggests that the improvements happened as a result of the incentives rather than of the purchase of new materials or changes in practices that were paid for with the additional resources. In other words, it seems that schools worked to improve their results in order to become eligible for the new performance-based grant. However, after the program was implemented, the new funding had little effect on student outcomes.

Overall, the jury is still out on whether adding conditionality (for example, restricting eligibility to top performing schools or distributing funds based on a series of performance indicators) to a grant increases its effectiveness. It seems that both context and design can be critical for the effectiveness of school grants. This approach worked in Senegal (perhaps because the grants were spent on human capital rather than materials), but it had mixed effects in Indonesia. Perhaps the effectiveness is driven by other design variables.

### Design Issues: Equity

An additional issue to consider regarding performance-based grants is equity. If grants are conditional on performance, the risk is that those schools that are already doing quite well will receive even more money. Higher-performing schools tend to have more affluent students since higher income is heavily correlated with good educational outcomes. This is an issue worth mitigating because otherwise financing can become regressive. In developed countries, a classic example is the No Child Left Behind reforms in the United States.[97] Since accountability was based on proficiency scores, it disproportionately penalized schools in low-income areas.

In the Indonesia grant program financed by REACH, measures were included to mitigate these equity concerns. For example, schools competed against other schools in their own districts rather than nationally. Also, one of the metrics used to calculate which were the top performing schools was absolute change in performance, which in effect benefited poorer performing schools. Since schools from the same district generally have similar socioeconomic profiles, this reduced the amount of inequity in the final allocation. However, it was not enough to eliminate inequity, and higher performing schools were still more likely on average to receive the performance grant. They were also more likely to have improved scores than the schools at the bottom of the distribution.

Ideally, a performance-based grant intervention should not only incentivize the entire distribution of schools to improve their learning outcomes but also close the gap between the best and worst performing schools. However, there is no guarantee that this will happen. In the Indonesia REACH grant, there were heterogeneous effects. Among primary schools, equity increased because the worst performing schools improved by more than the better performing schools. However, among junior schools the opposite was the case. Another way to mitigate inequality between schools would be to modify

96 Al-Samarrai et al (2017)

97 Kim and Sunderman (2005)

the allocation formula. For example, the formula could specify that only schools in the bottom socioeconomic quartile are eligible to participate. Alternatively, the weighting of each component in the allocation formula could be adapted to benefit disadvantaged schools.[98]

### Design Issues: Household response

Another design issue mentioned by the literature is how households react to the creation of a school grant program. One evaluation looked at whether it makes a difference when schools receive unannounced grants as opposed to grants that are expected, and found that it does.[99] When households expect the grant, they offset it by reducing their own spending on education. If the grant is unexpected, they do not reduce their spending. Using data from Andhra Pradesh in India and from Zambia, the authors found that unannounced grants led to improvements in student learning but announced grants did not. In the case of India, this amounted to improvements of 0.08 and 0.09 SD in language and mathematics test scores respectively (for a grant of US$3 per student). In Zambia, the government assigned block grants that also cost a little under US$3 per student. The evaluation found that test scores in language and mathematics both increased by 0.10 SD.

Of course, this does not mean policymakers should design unanticipated grants. That would be impossible. However, the authors suggest two alternatives. First, providing a larger grant (since the grants provided under both programs were small) that makes it impossible for households to offset. And second, focusing on providing resources for inputs that are harder to substitute.[100] The authors mention extra teachers or infrastructure, but funds for capacity building or other combined interventions would also qualify. There is a growing amount of evidence (which we will review in the following section) on the benefits of targeting several constraints at the same time.

### Design Issues: Long term effects

As mentioned, performance-based school grants are often design to incentivize improvements in management and allocation of school inputs. Could they have a long-term effect on outcomes?

Unfortunately there is not much available research in education, but in healthcare RBF there is some evidence that short-term transfers to frontline agencies can lead to long-term changes in behavior. For example, the temporary subsidies to health clinics under Plan Nacer in Argentina led to permanent increases in the provision of prenatal care and healthcare packages.[101] The implication is that the clinics did not provide these services before the program not because of the early fixed costs of adopting them but rather because of their perception that they would yield low returns. In other words, the incentives helped to overcome organizational inertia to reach a new and better equilibrium in healthcare delivery.

In the Argentine case, subsidies to clinics helped to overcome an organizational coordination problem. In other RBF areas such as teacher incentives or cash transfers, constraints to improve outcomes are not organizational, so the effect could be more limited. But in schools organizational constraints can be a limiting factor. In that case grants could—perhaps by helping the school move to a better management equilibrium—lead to long-term effects.

### Conclusion

Unfortunately, the evidence base on the effectiveness of performance grants is still quite limited. As the number of interventions

98 Al-Samarrai et al (2017)

99 Das et al (2013)

100 Das et al (2013)

101 Celhay et al (2015)

increase, it might be possible to tease out more factors that determine whether a program is successful or not. For now, it seems that in some cases they can work (as in Senegal where the school performance grant led to improved outcomes), but in others there have been more mixed results (like the REACH-funded evaluation in Indonesia, which saw improvements in junior secondary school learning outcomes but no improvements in primary schools). Often additional money is a necessary but not sufficient condition for schools to improve learning outcomes, especially if households adjust their own spending in response. A promising research agenda is the combination of school grants and other RBF interventions, which will be discussed in the next section.

## Combining RBF Interventions to Overcome Constraints

### Targeting different levels

One of the lessons that we learned from the evidence base is that interventions tend to work better when they are combined. An emerging literature seems to suggest that targeting RBF interventions at different levels may have strong synergies.

For example, a pay-for-performance scheme in rural Uganda raised attendance rates and improved student learning outcomes but only when complementary inputs were also provided, in this case textbooks.[102] Student attendance rose by around 0.56 to 0.60 SD two years after its creation, but gains were driven by schools that had access to textbooks. The evaluation shows that most teachers in the intervention increased their levels of effort, but that effort was only transformed into improved learning when the textbooks were available. In these schools, students had higher scores (0.11 SD) on the exam questions that were related to the topics and content covered in the textbooks.

Similarly, an intervention in Mexico that targeted several RBF levels found that incentives that only affected teachers were not as effective as incentives that covered teachers, principals, and students.[103] This shows the importance of making sure that the design of the intervention aligns the incentives of all agents involved. Finally, a recent intervention in Tanzania tested a program that provided a cash grant for schools, a pay-for-performance scheme for teachers, or both together.[104] The cash grant and pay-for-performance schemes had no significant effect separately but combined they increased test scores by an average of 0.12 SD.

### Combining RBF and institutional capacity building

RBF interventions can also be combined with capacity building. Some evidence suggests this can lead to good results. For instance, stand-alone grants seem to be less effective than in combination with other interventions.[105] This may be the case because schools are not aware of the best ways to improve their learning outcomes. There is some evidence from the management literature that firms and bureaucracies do not adopt good management practices automatically but can benefit from them when they are exposed to them.[106] Therefore, providing financial resources is not enough to overcome resistance to adopting new practices. Often, interventions that provide complementary inputs (such as training, follow up visits, or other capacity building) are required.

The evaluation of the Indonesia grant found some of the same issues. Schools were found to invest their grant money in hiring fewer contract teachers and spending more on

102 Gilligan et al (2018)

103 Behrman et al (2015) cited in Murnane and Ganimian (2014)

104 Mbiti et al (2018)

105 Das et al (2013)

106 Bloom et al (2013)

inputs that are not correlated with improved outcomes, such as school infrastructure. In Tanzania and Kenya, many principals had little knowledge of the specifics of the grants that their schools received, with 60 percent of Tanzanian principals not knowing how much money they were eligible to receive and 35 percent of their Kenyan counterparts not knowing the size of the grant for non-teaching expenses.[107] This suggests that there is space to include guidance for principals and/or other capacity-building in such programs.[108]

For example, another intervention in Indonesia evaluated different combinations of grants, training, community participation, and elections to school boards.[109] While the grants on their own and the grants plus training had no effect, the grants plus community links (in this case, involving the village council in the planning meetings of the school committee) led to improvements in learning outcomes. It increased test scores in language by 0.17 SD.[110]

This is illustrative of the importance of complementarities between types of interventions and the sensitivity of interventions to small changes. A recent study looked at two versions of a literacy program in Northern Uganda.[111] One was the original program, and the second kept most of the features of the original program but with small changes that reduced the costs by 60 percent to make it scalable (such as removing some of the expensive materials). Whereas the original program had a very sizeable impact (0.64 SD in reading and 0.45 SD in writing, some of the largest reported in the literature), the low-cost program had no significant impact on reading and a large negative impact on writing (-0.3 SD). Further analysis suggested that these differences were due to large complementarities between inputs in the original program, such as teacher quality and materials.

The lesson to be drawn with regard to the design of RBF interventions is that careful thought must be given to every step in the process from financing to results. How are inputs going to interact with each other? Will the funding provided be enough for schools/teachers/students and families to accomplish what is required to improve learning outcomes? And if the answer is no, what other activities can be included that could enhance the effects of the RBF intervention?

107 Mbiti (2016)

108 Al-Samarrai et al (2017)

109 Pradhan et al (2014)

110 A fourth treatment that included grants plus community links plus elections to the school board improved test scores even more - 0.23 SD for language.

111 Kerwin and Thornton (2018)

## Summary

While the research on RBF and teachers, RBF and students and families, and RBF and schools is not comprehensive, there is substantial evidence to suggest the following conclusions:

1. Teacher incentives can but do not always improve teacher attendance and student learning. The design of the incentive scheme and the context matter. The effects are larger and more positive in developing country contexts.
2. Student and family incentives (such as CCTs, for instance) can reduce school dropout and increase school attendance, though the evidence for its effects on student learning are more mixed. Conditional transfers to students tied to their own learning are a promising area of future research.
3. The evidence on performance-based grants is still quite limited. For now, it seems that in some cases they can work, especially when grants are combined with other interventions such as capacity building (for example, to principals and school committees) or when money is spent on inputs that affect learning outcomes.
4. There is growing evidence that combining different RBF interventions within the same program can generate better results than using any one intervention alone.

# Part II – RBF and Governments: Making RBF More Effective

**When it comes to RBF and governments, few standardized studies are available,** which makes it difficult to make definitive or comprehensive statements about how RBF can be more effective in development projects.

However, as the number of RBF projects grows in the education sector, there is more operational experience from which to learn. This discussion is structured around the project cycle (see Figure 2 below) because the lessons differ for each of the stages of the cycle. The information in this part of the report is taken directly from our qualitative survey data that reflect the experiences of development agency staff in the field, along with examples from project documentation and academic research.

**In this section, we aim to highlight the practical experiences of using RBF with governments,** and the information that we present mostly relates to results-based financing agreements between a donor and a country client.

*Figure 2: A Typical Project Cycle*

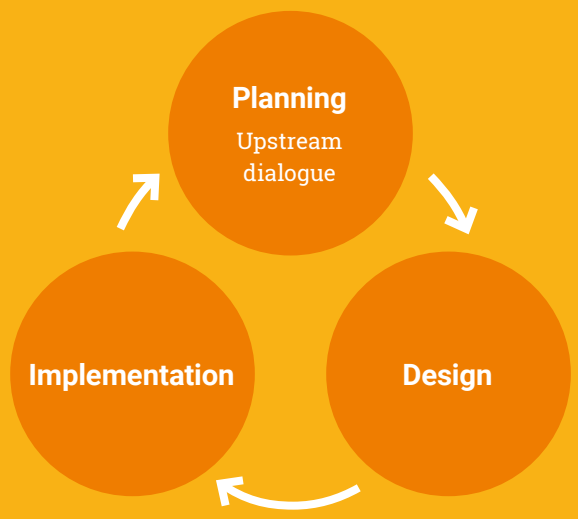
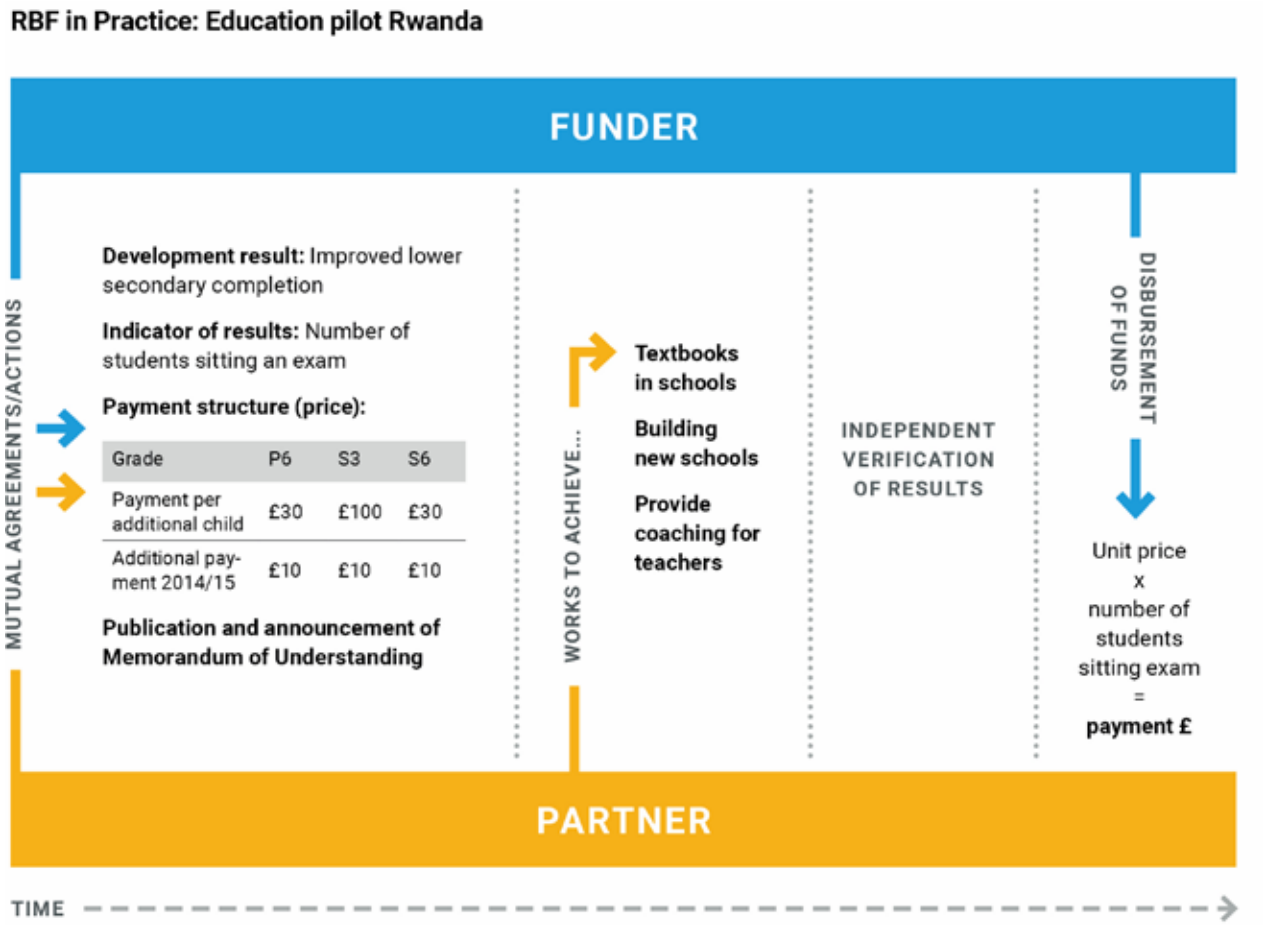


*Figure 3: How RBF and Governments Work*

This figure shows how the relationship between RBF and a national government typically functions. The funder is the donor, and the partner is the country client. The donor and client must mutually agree on the expected results of the intervention, the indicators that will be used to measure those results, and what values the indicators must reach. The client then must work to achieve the indicators. Once these results have been verified, the donor then disburses funds to the client.




*Source:* DFID (2014)

There are very few education projects at the national-level that have closed and been independently evaluated, and thus, little rigorous evidence exists of its effectiveness relative to other development financing. For example, there is some suggestive evidence that RBF may be more effective than other financing modalities in health, but more research is needed.[112] Looking at eight RBF projects across three sectors, a recent paper claims that there is no evidence that RBF projects lead to fundamentally more innovation or autonomy,[113] though those factors may not be the primary benefits of RBF. In education, the evidence is even more limited, and researchers have noted a lack of documentation of practical, real-life experiences with using RBF in comparison with other social sectors such as health.[114]

112 Grittner (2013)

113 Clist (2018)

114 R4D (2016)

In general, it is difficult to ascertain the direct impact of RBF in comparison to other financing given that it is rarely used in isolation.[115] Another challenge is that there is usually no counterfactual situation with which to compare it. There are no known experiments where researchers have compared a situation with RBF to a situation without RBF as it would be difficult to create conditions under which the two situations would be comparable. It is also just as challenging to know whether countries would have funded similar activities to achieve results using less money. Furthermore, a typical project that uses RBF may target a broad set of indicators, some of which are tied to financing and some of which are not, and this can make it difficult to ascertain the overall effects of RBF within the larger project.[116] Despite these challenges, some lessons and best practices have emerged related to how and when RBF can work. In order to promote the use of diverse and flexible approaches to solving complex education development problems, donors and clients require a range of financing options, with RBF being one potential choice.

## Choosing RBF: Commitment, Cautions, Cost, Context

Table 7 below shows the four considerations that must be borne in mind when selecting which type of RBF to use in any given project — commitment, cautions (risks), cost, and context.

*Table 7: The Four Cs*

| CONSIDERATIONS TO BEAR IN MIND WHEN CHOOSING RBF |
|---|
| **Commitment** |
| **Cautions** |
| **Cost** |
| **Context** (country systems; capacity; conflict, fragility and violence) |

115 UNESCO (2018)

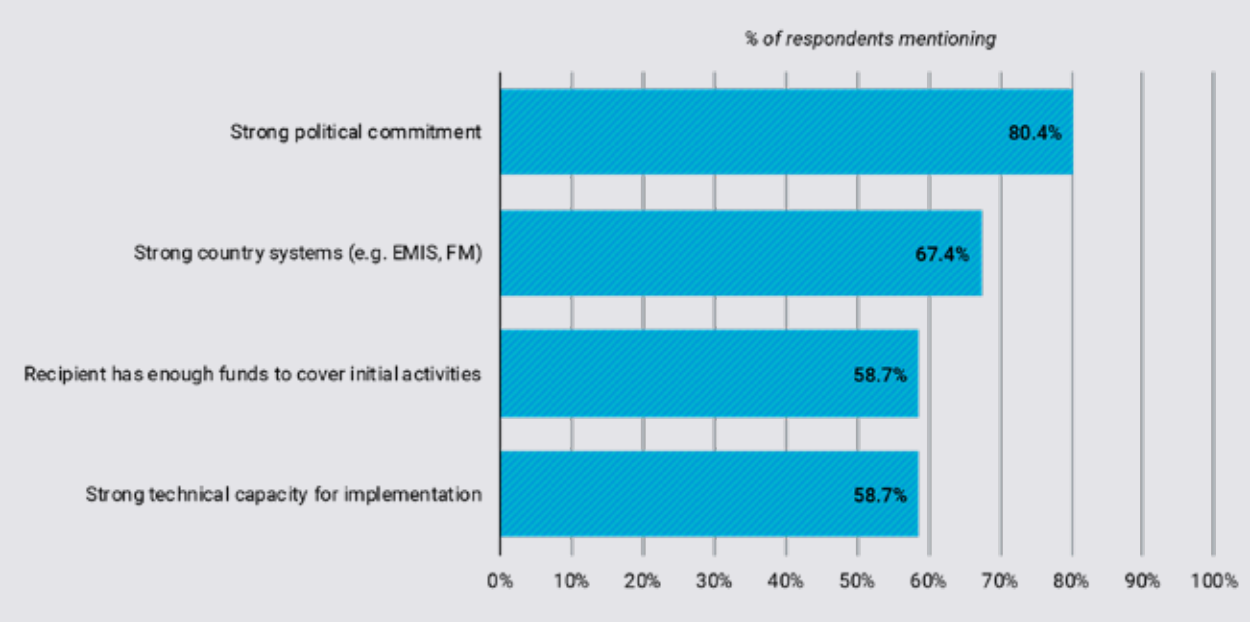
116 Grittner (2013)

## Commitment

At the project planning stage, the most important thing to consider when choosing RBF is whether there is mutual agreement between both parties to use it. While this applies to other financing modalities as well, it is often discounted. For RBF in particular, it is an important signal of political commitment, especially given that there has been criticism of RBF as a new form of conditionality and as a tool that donors use to ensure that the recipient's incentives are aligned with theirs.[117]

Our survey responses also point to the importance of political commitment. One respondent wrote, "I think we can work around the financing, weak systems, weak capacity. I mean, it's not ideal. But it can be accounted for. We can't, however, work around the lack of political will." In fact, strong political commitment was the most commonly mentioned factor needed for RBF to be successful (see Figure 4).

***Figure 4: What Conditions Are Necessary for RBF to be Successful?***




While political commitment and ownership by policymakers in the recipient country are critical to the success of all development projects, it is also important to acknowledge the inherent power imbalance that exists between donors and recipients. Oftentimes, countries are not in a position to refuse funding, and an added complexity with results-based financing is that country governments may not fully understand how the modality works (this will be further discussed in a later section).

Ultimately, both parties have specific interests. These can range from the seemingly innocuous to the more problematic. One argument is that donors use RBF to try to make aid more efficient and to get the results they want, without ensuring a mutual agreement with the recipient government on which results will be linked to financing, without putting any effort into explaining how RBF works, or providing the recipient country with the support needed to actually achieve those results. On the recipient's side, the fear of not receiving funds

117 Clist (2016)

may lead policymakers to choose easy-to-achieve targets in order to ensure that they receive the payments.

Regardless, RBF has the best chance at success when both parties are equally committed to it and understand the risks involved. Here are two examples of when the interests of the parties are and are not aligned.

***Recipient Feels "Pressured" into Accepting RBF***

Based on survey feedback, there has been some indication that RBF is being heavily championed by a number of development agencies. Roughly 25 percent of respondents indicated that their agency's position on RBF was positive, with some even saying it was "hyper-positive" or "positive, perhaps excessively so." One respondent gave an example of a middle-income country in the Middle East and North Africa region where RBF was chosen as the financing modality before the project's objectives and activities were fully identified. The project manager from the Ministry of Education indicated that, while the government wanted to achieve results, it would have been better if they could have introduced some of the necessary reforms to strengthen their own country systems prior to implementing RBF. This project is still ongoing, so it is unclear whether RBF will be a success or not, though the government's commitment is now there. There may also be other country context issues irrespective of RBF that will generate political instability, which may alter the country's ability to achieve some of the indicators.

However, while there is some qualitative evidence that development partners are choosing RBF as a financing instrument without much in-depth consultation with country governments, our survey results show that this does not seem to be the overarching pattern as is illustrated in the next example.

***Recipient is Keen to Pursue RBF and Has the Necessary Political Will***

In 2008 a new government in Pakistan came into power with a strong commitment to the World Bank-financed Sindh Education Program and with the desire for World Bank's assistance in further refining the program's focus on results. Given that Pakistan was in the process of decentralizing responsibilities in order to improve public service delivery, the government also wanted to institutionalize results-based budgeting. For this to work, they needed to introduce RBF at different levels and agreed to a series of disbursement-linked indicators that were meant to reinforce the priority areas of the program. RBF was effective in this instance because of the sector-wide approach that was taken, which required "strong political commitment and ownership (which) is critical for… addressing governance constraints to effective service delivery."[118] In addition, the use of RBF complemented the support being given by other development partners. At that time, the main donor was the European Commission, which has some elements of RBF in its budget support model. The success of the first Sindh Education project led to a second iteration, which also used RBF.

118 World Bank (2009)

The results-based design of the original project "likely helped in orienting and focusing the Sindh government's efforts on agreed program implementation progress and performance targets. In particular, the disbursement-linked indicators (DLIs) likely helped to promote and protect the continuity of politically difficult, governance-oriented reforms undertaken by the Sindh government."[119]

One of the most critical aspects of RBF is the need to communicate to the recipient upfront how RBF works, whether the recipient is a national agency, a sub-national agency, and/or a direct service provider. Any group of individuals who will be affected by RBF should be made very aware of how the RBF scheme will work. In two DFID-sponsored RBF schemes, this lesson proved doubly true—the Girls Education Challenge and a pilot RBF intervention in Ethiopia where RBF was found not to have any discernible effect.[120] The evaluation team noted that the project had not been effective in disseminating information about RBF to the regions where, even after two years, few officials, including head teachers, had heard of the pilot.[121]

Many of the existing studies of RBF examine the relationship between the donor and the recipient (the government of the recipient country, usually the Ministry of Finance), but in reality, governments who are genuinely interested in results can also use it internally at all sub-national levels.[122] For example, there is research underway in Morocco, Sudan, and the Dominican Republic that is looking at how performance contracts between national and sub-national government levels might improve education quality.

### Costs and Benefits (Advantages and Disadvantages of RBF Over Traditional Aid)

The research that outlines the potential advantages of RBF generally makes two arguments in its favor: (i) it demonstrates the impact of aid money and (ii) it is more effective than other forms of aid.[123] One study in particular has argued that RBF can help to maximize the alignment of interests between donors and client countries, among other things, which can make aid more efficient.[124]

While these theories may be true in some form, over half of respondents (57 percent) to our survey indicated that RBF helped the recipients to "achieve results that were previously not achieved through other financing modalities." This sentiment was confirmed by another survey question that asked respondents what the biggest benefit of RBF was over the financing of inputs. The overwhelming majority, 96 percent, indicated that RBF produced a "sharper focus on results." The second biggest benefit that they identified (64 percent) was that RBF "relies on and/or strengthens country systems." However, while the survey takers were able to see the promise of RBF, they also acknowledged that it often is not used very effectively (43 percent) and requires more implementation support to ensure good results. The challenges related to implementing RBF are discussed in more detail in a later section.

119 World Bank (2013)

120 Coffey (2016) and Cambridge Education (2015)

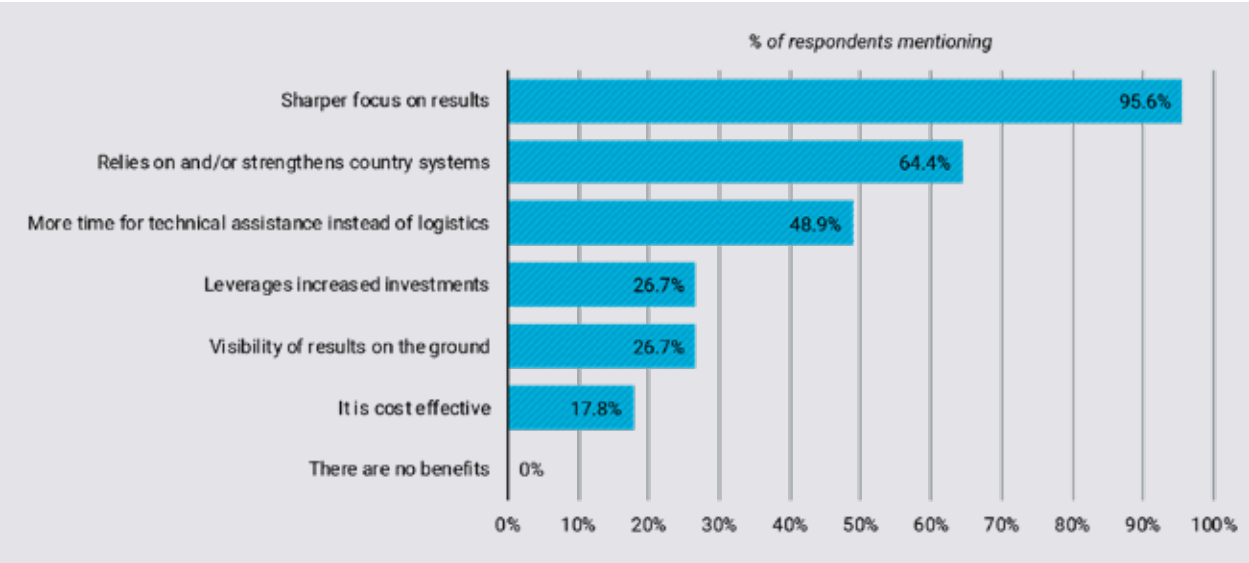
121 Cambridge Education (2015) and Coffey (2016)

122 Birdsall and Savedoff (2012) and Clist and Dercon (2014)

123 Clist (2018)

124 Clist and Verschoor (2014)

*Figure 5: What Are the Biggest Benefits of RBF over Financing Inputs to Achieve Results?*




Although RBF has become a mainstream financing modality and can help donors to encourage a greater focus on results, some interviewees felt that clients themselves may also be keen to do the same. One survey taker wrote, "I first was on the implementation side, based within the Ministry of Education in Ghana, and now I'm on the donor side (World Bank) for the same project, which is an RBF, secondary education project. I found that, due to the emphasis on results, my government colleagues took much more ownership over the project that some of the other donor-funded projects in Ghana."

While there are strong proponents of RBF, there are also detractors, those who believe that identifying good indicators is difficult, that it undermines country ownership ("new conditionality") and that it does not produce good value for money.[125]

***Cost-effectiveness of RBF***

Although RBF can be used to get both donors and recipients to focus more carefully on results, do these results come at a higher price than using other forms of aid? To date, there is no consensus on whether RBF has a clear cost advantage over traditional financing.[126] In this report, cost-effectiveness is defined as the ability of RBF to produce results more efficiently and effectively than traditional financing. Measuring the cost-effectiveness of RBF is still problematic because of the limited existing evidence base. While there have been a number of RBF interventions in education, they are rarely implemented alongside other modalities, making it hard to compare RBF's relative value for money. Also, there are no comprehensive cost-effectiveness frameworks for RBF in education that would enable these comparisons. In other sectors, such as health, toolkits have been developed that provide guidance on how to evaluate the cost-effectiveness of RBF in comparison with other aid modalities.[127]

125 UNESCO (2018)

126 Paul et al (2018)

127 See for example, Shepard et al (2015).

The Center for Global Development has argued that choosing a lending instrument like PforR does not create any financing additionality because countries can still avail themselves of other traditional lending instruments.[128] The IDB looked at whether RBF (donor to country client) was more effective than traditional aid in the health sector in El Salvador and found that RBF led to higher growth in many of the indicators measured. In municipalities receiving RBF, preventive visits increased by 42 percent compared to a 20.9 percent increase in traditional aid visits. There was a more modest difference in increases in outpatient visits (6.7 percent in RBF villages versus 4.2 percent in traditional aid villages).[129] However, these improvements seemed to be due to an expansion of infrastructure and increased medical staff rather than to divestment from other areas.

In education, the idea that RBF should achieve greater "value for money" comes off particularly strongly in the evaluations done of some of the DFID's early investments using RBF. An evaluation of the use of RBF in the Rwandan education sector found that increases in completion rates (for primary, lower secondary, and upper secondary education) during the implementation of the RBF program could not be attributed to the program itself and instead were the result of other factors.[130] However, the evaluation did show that investing in the Rwandan education system was sound and good value for money regardless of which aid modality was used because the benefits of increasing access, retention and completion clearly outweigh the costs.[131] It also suggested that the value for money of the RBF program would have been greater than that of other financing modalities. Therefore, it seems that implementing RBF would not have come at a higher price than implementing other financing modalities, at least theoretically.

The case of Ethiopia is somewhat similar. An impact evaluation found that the program had had negligible effects on the outcome (in this case the number of students taking and passing an exam).[132] Therefore, it is not possible to say whether the program was good value for money or not. Nonetheless, the evaluation found it to be a relatively low-cost alternative since it had low transaction costs and did not disburse funds if there were no results.

Despite the possibility that RBF may increase the cost-effectiveness of education financing, only 17.8 percent of respondents from the agencies in our survey (including DFID) indicated that cost-effectiveness was one of the biggest benefits of RBF. This may be because those responding to the survey were not directly responsible for discussing the modality with the country government or that the loan amount had already been determined or agreed upon and RBF had simply been chosen as the way for the funds to flow. Alternatively, many of those surveyed may not have taken cost-effectiveness into consideration as a reason to choose or not choose RBF. It should be noted that many of the respondents had not seen a project through to completion so may not have assessed its final costs.

There is some evidence outside of the education sector that the initial costs of acclimating a country to an RBF approach may be higher than those needed for traditional financing. In Ethiopia, the first World Bank-funded PforR operation was incredibly difficult and costly to prepare and was not advantageous to the Bank on cost grounds. The evaluation of the project recommended that the World Bank invest sufficient resources upfront in

128 Gelb et al (2016)

129 Bernal et al (2018)

130 Upper Quartile (2014 and 2015) and Cambridge Education (2015)

131 Upper Quartile (2015)

132 Cambridge Education (2015)

future to ensure that teams have the capacity to explain RBF (in this instance, the PforR instrument) well enough so that the country can make an informed decision about the financing modality.[133] A more recent study has found evidence that RBF is more expensive and not necessarily more efficient than traditional financing, mostly due to the costs of supervision and independent verification. For example, in a World Bank-supported health project in Benin, for each US$1 paid to providers, half (US$.50) was spent on verification.[134]

### Cautions

Another reason why countries (and organizations) are skittish about using results-based financing is that they have to assume more risk.[135] If they fail to achieve the required results, they will not receive any money. Also, some countries may not have enough upfront financing available to cover the costs of achieving results. These are legitimate fears. An evaluation of the Girls Education Challenge (GEC), a DFID-funded initiative that seeks to improve learning amongst the poorest girls, found that many organizations could not bear those risks.[136] Although the GEC is not technically an RBF and government scheme, it is used here for illustrative purposes because it still involves donor funding, except that the donor funds are directly channeled to service providers.

According to the results of our survey, over half (54.3 percent) of respondents indicated that when the expected results were not met in the projects that they managed, they did indeed withhold funds. This is a politically difficult choice to make, which may also explain why the remaining half of respondents indicated that they scale back the indicators when the recipient does not achieve the required results. Other write-in responses alluded to similar actions, including "watering down" indicators or extending the project. The failure to achieve targets will be further discussed in the implementation section.

The liquidity constraints faced by country governments and other incentivized actors can be mitigated through such flexible measures as providing advances to cover initial expenses and/or staggering the indicators, in other words, setting achievable targets at first and building up to more difficult ones as the project progresses. This was done in the Sindh education project mentioned earlier and will be further explored later in the section on indicators. Recent practitioner experience reflected in informal World Bank guidance has shown that it can be helpful to use "zero DLIs" — those that are easier to achieve and thus to earn funds — as a way to speed up implementation as well as to "boost morale and momentum."[137]

In addition to non-payment and liquidity concerns, there is also political risk involved in using RBF, particularly the risk that RBF can be more difficult to control when there are many political actors involved. This can complicate accountability and coordination, for example, if the agency receiving the money is not the agency in charge of achieving the results. In some cases, the agency in charge of achieving the results (usually a line ministry) might wonder why they should prioritize those results since they are not the ultimate recipients of the funds (usually the Ministry of Finance). One respondent to our survey wrote about a social protection project in Nepal where there was some question over whether RBF should be used as the team leader had doubts that the Ministry of Finance (the funds

133 IEG (2016)

134 Paul et al (2018)

135 UNESCO (2018) and Results for Development (2016)

136 Bond (2017)

137 Sabarwal et al (2016)

recipient) would be capable of holding the line ministry (Ministry of Education) accountable for achieving the results. In the end, RBF was still chosen as the financing modality because other government actors were also involved, but the team was cognizant that "successful implementation depended on how pro-actively [they] engaged these other institutional actors beyond the immediate counterpart agency. The jury is still out as to whether this will work as intended."

With these examples in mind, both types of risks should be assessed in the planning stage of a project. While traditional input-based financing can involve similar risks to RBF, the stakes are often higher for recipients because of the threat that they will not receive the funds if they do not achieve their targets.

Another risk that underscores the importance of explaining the details of RBF upfront is when countries or organizations unwillingly or unwittingly agree to RBF without fully understanding what they are signing up for. A good example of this is the GEC. In a process evaluation, the evaluators indicated that, although the project team mentioned RBF upfront in all of its guidance to service providers, their understanding of it evolved over time, and DFID was unable to issue specific guidance about it, due to the lack of consistent definitions within the organization. This meant that, while many applicants signed on to the concept, they did not really know what it would entail. This caused confusion and frustration among applicants as well as delays in the implementation of the GEC.[138] The evaluators subsequently found that it would have been more useful to develop guidance on RBF and begin disseminating it as early as possible so that applicants could make informed decisions about whether they wanted to apply to be a part of the GEC given the RBF requirements. Overall, the evaluation did not recommend using RBF as the only financing modality for the GEC.

If a country does not have the ability to mitigate risk or if the project guidelines and procedures do not allow for sufficient mitigation, then it may be necessary to rethink the use of RBF in that particular country.

## Context

RBF, like all financing modalities, needs to be designed to fit the specific context within the recipient country, and there is no "one size fits all" design. However, some contexts may be more conducive to the successful implementation of RBF than others. The key criteria that are likely to lead to its successful implementation are: (i) the pre-existing involvement of the development agency in the sector or in a government program and (ii) the pre-existence of strong financial management systems and EMIS in the country in question.[139] Other important factors that may influence whether RBF is the appropriate modality are the capacity of the state and whether the country is affected by conflict, fragility, and violence.

***Country Systems***

According to our survey data, 67.4 percent of respondents believed that the second most necessary condition for RBF to work was the existence of strong country systems, the first (as has already been noted) being political will (80.4 percent). Two of the key country systems needed for RBF to be implemented successfully are financial management (FM) systems and an education management information system (EMIS). Financial management systems are necessary to ensure that funds are well managed, while a functioning EMIS is needed for all monitoring and evaluation efforts, which are the key feature of RBF operations.

An example of an intervention that benefited from all of these prerequisites being met is the

138 Coffey (2016)

139 ADB (2017)

Asian Development Bank's Additional Skills Acquisition Program in Kerala, India. This program is a post-basic education results-based financing operation that was designed as such because all of the necessary country systems were in place. The RBF approach was chosen so that the implementing agency would have the freedom to make any required changes in real-time, subject to the proviso that the ultimate results were met.[140]

The project used the government's financial management systems for its budgeting, accounting, reporting, monitoring, and auditing arrangements. According to the National Institute of Public Finance and Policy, India scored very highly on the public financial management dimension of "comprehensiveness and transparency."[141]

In addition, the country's existing management information system (MIS) was able to provide program managers with critical information for program planning such as gender, inclusiveness (specifically of socially and economically marginalized or differently abled students), geographical spread of students, and sector training. The MIS was already set up to track output and outcome indicators, including the disbursement-linked indicators. The MIS also facilitated evidence-based planning and could flag potential problems with the program early on.

Although all of the requisite country systems were functioning and available to be used in the Kerala project, the project documentation also noted, "While country systems (especially data systems) are critical, RBF can be designed to strengthen those systems and… can provide advances and strengthen technical capacity." This emphasizes the importance of using monitoring and evaluation as a feedback mechanism, which will be further detailed in the implementation section.

***Capacity***

RBF can work in a variety of contexts, but the capacity of the implementing country is key. Capacity, in this report, is defined as the ability of a country government to implement and monitor RBF activities or at least to have the desire to build its capacity in order to be able to do those two things.

It is important to dispel the implicit assumption that low capacity means lower income. This is not always the case, given that there are some low-income countries that have the capacity to implement and monitor RBF, such as Tanzania and Rwanda.[142]

For example, Rwanda is often listed as a forerunner in the use of results-based financing, especially in the health sector. Their experience with the modality began in the early 2000s when donor support for reconstruction after the 1994 genocide waned and health facilities were once again reliant on user fees to keep going. In addition, health workers were poorly compensated and thus not keen to work in the public sector. To refocus efforts on increasing use and coverage of health services and improving their quality, results-based financing activities were initiated, which were designed to increase the use of services by incentivizing health providers. The initial schemes were very successful and were eventually scaled up and tested in other provinces. For example, some research has suggested that it helped to increase measles immunization by 11 percentage points (compared to only a 1 percentage point increase in the non-RBF areas).[143] The quality of service delivery also improved, with RBF areas scoring 73 percent in a composite quality score versus 47 percent for non-RBF areas. Ultimately, despite Rwanda's history of conflict and violence, the government has remained

140 ADB (2014)

141 ADB (2014)

142 See examples in Andrews et al (2017).

143 Rusa et al (2009)

committed to using results-based financing.[144] As previously discussed, over the past decade it has piloted initiatives in the education sector, though with less impressive results.[145]

According to the evaluations of the Rwandan RBF activities, they succeeded for several reasons. First, they were built on three existing donor-funded pilots, which facilitated the scaling up process. Lessons were learned from the pilots on key issues such as the need for robust information systems or the best way to manage and distribute the funds, and these were put into practice in the scaled-up nationwide program. This echoes the conclusions reached in other studies of RBF programs about the importance of experimentation and context in their implementation.[146] Second, they benefited from broad political leadership and political will at the highest level in Rwanda, which evidence suggests is key for any reforms to be successful.[147] Third, the Rwandan government had already demonstrated that it had the capacity to responsibly manage funds and monitor indicators, which led them to promote institutional capacity, particularly for service provider contract management, into the public system rather than rely on donor accountability.[148] The Rwandan case shows how even very low-income and fragile countries can successfully implement RBF (Rwanda has a per capita GDP of around US$750).[149] Also part of Rwanda's context is the government's development of the Imihigo system, a "Home Grown Solution", which is based on contracts between national and subnational levels.[150] This type of government commitment resembles what was mentioned earlier in the section on *Choosing RBF*.

Another dimension of the capacity question is that capacity is inextricably linked to country context and project design. One respondent to our survey wrote, "I've worked on [RBF projects] in relatively high capacity contexts (a reformist state in Brazil) and very low capacity contexts (Nepal). The country I am currently assigned to is an extremely fragile state with ongoing conflict, a fragmented state, and endemic corruption. I wouldn't categorically deny the applicability of RBF based on broad categorization of 'contexts' but I do believe that in each place it really is important to think through why an incentive-based approach is superior to traditional input financing, what are the institutional (or fiscal) pre-conditions for this to work, and whether enough of such pre-conditions are in place."

These examples demonstrate that it is useful for the designers of RBF projects to identify upfront: (i) the kind of capacity that will be required to implement the project and (ii) whether this capacity (both political and technical) exists in the intervention. The type of capacity required varies and can be straightforward, as in Rwanda, or more complex, as in pay-for-performance schemes for teachers.[151] Similarly, schemes that will completely revamp the existing system will require more political leadership than others that use existing structures. As to whether capacity exists to implement the specific

144 Rusa et al (2009) and Rusa and Fritsche (2007)

145 Upper Quartile (2015)

146 See Andrews et al (2013).

147 Andrews (2013) and Andrews et al (2017)

148 Rusa et al (2009)

149 World Bank (2018)

150 Klingebiel et al (2016)

151 As suggested by Barleavy and Neal (2012).

intervention, in Rwanda, RBF was built onto existing programs that had already been proved to be effective. In other words, as long as there are pockets of effectiveness where the required capacity exists (or willingness by the government or actors to invest in creating them), it seems that RBF can be implemented successfully.[152]

***Conflict, Fragility, and Violence***

From our survey data, it is clear that the vast majority of respondents (79 percent) felt that RBF could be introduced in fragile and conflict/violence affected (FCV) areas. Even if this is the case, RBF in FCV contexts generally requires more customization. In SIDA's internal guidelines, there is a section dedicated solely to "special design considerations in fragile states," which rightly indicates that, thus far, the experience is limited and not very conclusive.[153] However, the guidelines note that there has been more experience in the health sector with operating RBF programs in FCV contexts, and the general approach has been to pilot RBF in a particular region or province and then scale up based on the success of the pilot. This was precisely the *modus operandi* in the successful health intervention in Rwanda that was discussed in the previous section.

In these settings, RBF might be better deployed to incentivize system-building than outcomes at first, as these systems will make it possible to set intermediate and outcome-level indicators in the future. For example, in Lebanon, which is dealing with a Syrian refugee crisis, RBF is being used to move from a crisis situation to a more sustainable one by incentivizing the government to prioritize education quality for both Lebanese children and Syrian children. The project has nine indicators that must be met for funding to be disbursed, four of which are focused on "strengthened systems" such as improving data management, revising the curriculum, putting foundational policies in place, and increasing government capacity for planning and implementation. One survey respondent wrote, "I was skeptical that Lebanon would be able to successfully implement RBF, to bring more refugees into the systems. But it turns out I was wrong." Based on the most recent status report, there has been an increase in the enrollment rates of both Lebanese and non-Lebanese students.

The Dutch NGO Cordaid has been implementing RBF projects in FCV environments since 2001, particularly in the health sector. Cordaid introduced RBF to the health sector in Sub-Saharan Africa and has a few activities underway in the education sector. Cordaid believes that RBF works particularly well in FCV contexts because it allows for more flexibility in funding allocations for local health facilities to decide what is needed based on their neighborhoods' needs (this flexibility will be discussed further in the implementation section).[154] RBF can also be used to target specific populations. For instance, in the Democratic Republic of Congo, the number of safe deliveries of babies rose to 97 percent in RBF facilities compared to non-RBF facilities.[155]

Haiti is an example of a fragile, low-income country where the preconditions for the effective implementation of RBF were not in place. In particular, Haiti has been struck by numerous external shocks that have exacerbated the country's fragile state, including a devastating earthquake in 2010, a powerful hurricane in 2016, and long political transitions. These shocks have greatly diminished public sector capacity, particularly in the education sector. Haiti has

152 The idea of pockets of effectiveness has been discussed extensively in the literature, often under different names. See, for example, Andrews et al (2017) and Marsh et al (2004).

153 Olander and Högberg (2016).

154 See, for example, Results for Development (2016) or World Bank (2017) .

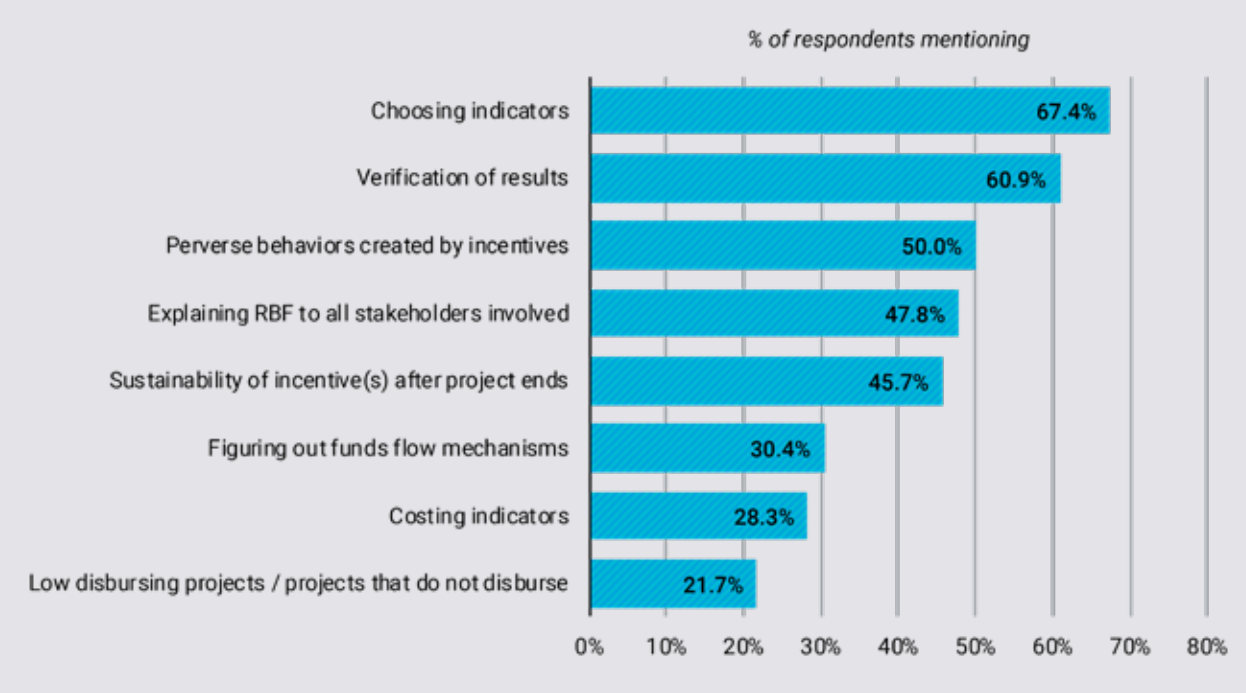
155 Cordaid (2017)

a unique education system in that the majority of providers are private (approximately 80 percent of students in the system attend non-public schools) and therefore are not within the jurisdiction of the Ministry of Education. To support access to education for the poorest children, the Government of Haiti has funded several tuition waiver programs where fees are directly paid to non-public schools on the condition that they meet a series of requirements related to education quality. Despite the good design of these programs, the government did not have a reliable set of indicators or monitoring systems in place to verify whether schools met the stated conditions. Thus, the government participated in an exercise to develop such systems in an effort to create a stronger link between data (indicators) and incentives. Haiti's development of a quality assurance system (QAS) to strengthen its capacity to implement RBF will be further explored below in the section on monitoring and evaluation.

## Design Priorities

While there is a general consensus in the development community that how a project is designed is critical for its successful implementation, it is not always clear which design elements are the most important, especially in RBF projects. In our survey, practitioners identified the two biggest challenges in project design as choosing indicators (67 percent) and verifying results (61 percent). These will be discussed in detail in this section, along with other design considerations that can lead to more effective RBF.

***Figure 6: What is the Biggest Challenge in Designing RBF Activities?***



### Cascading Incentives

In education, as in in health, there have been many RBF schemes that have targeted frontline providers, notably teachers and health workers. Interestingly, a very large majority of our survey respondents (90.9 percent) indicated that it was most important to incentivize national-level actors such as policymakers. In education, one of the challenges is that the main national-level actor is usually the Ministry of Finance, which often does not communicate to the Ministry of Education about the incentive scheme or does not cascade the incentive, i.e. keeps all of the disbursements at the Ministry of Finance level, or does not incentivize the Ministry of Education to achieve results that they are responsible for. The mechanics of how this relationship plays out is further explored in Box 1 on choosing financing mechanisms. Other than national actors, 61.4 percent of the survey respondents indicated that front-line providers (for example, teachers) were the most important people to incentivize, followed by schools (75 percent) and then meso-level officials (for example, district education officers) (84.1 percent). These responses may be skewed due to the targeted survey population (since staff at development agencies primarily interact with their counterparts at the central government level), but in the written responses to this question, respondents addressed this by indicating that teams "usually do not think through the potential cascading effects (or lack thereof) of incentives within the government structure." In one example, a respondent mentioned a pipeline project that had not addressed the disconnect between how donor financing would flow to which part of the government and who within the government structure needed to be incentivized to achieve the desired outcomes. Other survey takers indicated that, even though the donor community and development agencies are obviously not incentivized, they should also be held responsible for achieving results as well as the actors within the country. This idea of mutual responsibility for results is often overlooked in other theories of RBF where in-country agents are expected to innovate on their own.[156]

These responses show the importance of thinking through the results chain and how incentives can cascade down to every level. There is some operational evidence that incentives should be targeted to the responsible administrative levels where the action being incentivized is taking place, but this is not always the case.[157] In many projects where donors provide funding to incentivize country governments, other actors at lower levels are often bypassed or overlooked. This may not be intentional, but this is often the problem — that donors have not thought about which actors play a role at what stage of the results chain.

In the DFID-funded Ethiopia Secondary Education pilot, the national government was incentivized to increase the number of exam sitters, boys and girls, over the course of three years. The national government in turn passed down part of the incentive to the regional level, and within some regions, schools were responsible for spending the RBF funds, and some were also directly incentivized. Even though there was no short-term impact, there was some evidence that strategic thinking and prioritization improved at both the regional and school levels, though the evaluators indicated that their team did not visit enough schools in Year 3 to make a strong argument that the school-level changes were widespread.[158]

This example shows that the intervention's designers assumed that regions and schools would be influential in increasing the number of exam sitters, though it is unclear if that assumption is indeed true. The trickling down of incentives did create some other types of

156 Birdsall and Barder (2006)

157 Sabarwal et al (2016)

158 Cambridge Education (2015)

positive results, such as modifying the formula to reward only the most successful schools, supporting all schools to increase the numbers of both sitters and passers at the regional level, and supporting underperforming school to increase the numbers of both sitters and passers at the regional level.[159] Given that one flaw of the pilot was that most head teachers did not even know about the pilot intervention, it is unclear if providing incentives to other stakeholders such as parents and families and/or even students themselves would have made a bigger difference since presumably they may have had more control over whether they showed up for the exam. While there may have been other unintended consequences of that design, there is some evidence that those types of incentives can work to increase effort and attendance. While there is also some evidence that school-level incentives can produce similar

159 Cambridge Education (2015)

### Box 1: Choosing Financing Mechanisms for Cascading Incentives

Regardless of what result or stakeholder is incentivized and how much an indicator costs, there also needs to be a way to transfer funds. This is particularly important in RBF because the incentives need to be able to get to the right actors. Unfortunately, this aspect is not always carefully thought through during the design phase, which leads to problems in implementation. For example, in a higher education project in India, there was no mechanism for the central government to transfer funds directly to participating institutions, and thus the funds were first sent to state treasuries (World Bank, 2017b). The completion report noted that fund releases from the state treasuries to institutions took an inordinately long time — over 100 days in many cases and even 300 days in a couple of cases. Unfortunately, the central government did not have the ability to control or sanction those states that did not disburse funds, and delays in disbursement greatly affected the ability of certain institutions to comply with project milestones and diluted the effectiveness of RBF.

In a West African country, there were problems with funds flow that led to the cancellation of DLIs. The Ministry of Finance was not keen to disburse money to the Ministry of Education, which was responsible for achieving the DLIs. This was due to a number of factors, including an economic crisis in the country and the poor relationship between the Ministries of Finance and Education. This is not a unique situation. Especially when it comes to traditional aid relationships, in many countries the Ministry of Education is not the strongest line ministry, yet they are responsible for selecting the DLIs while the Ministry of Finance is the agency that receives the disbursements.

This type of issue has been managed through "results-based budgeting", which has been used in various projects. For example, the designers of the Jamaica Early Childhood Development Project included DLIs that required the Ministry of Education to prove that there was an adequate budget for achieving other DLIs for the same fiscal year and that the execution rate under the budget lines for DLIs exceeded 70 percent (World Bank, 2008). Many of the "budget" related DLIs in the World Bank DLI analysis (to be described in further detail in the next section) are tied to timely execution and/or adequate resource flows.

results, perhaps combining the two incentives would have created more of an impact, along with better dissemination of information about the RBF scheme. Without understanding the full Ethiopian context, it is difficult to know, but this example still shows that when designing a project, teams should question their assumptions and think about how incentives will trickle down and to whom and how those incentives might work together.

Some survey respondents argued that targeting the incentives to the wrong actors can have negative effects. If they are targeted to high levels (for example, the government), they have little hope of cascading, but if they are targeted to actors too low down the chain, the risk of perverse behavior grows as in the case of teacher performance pay schemes (which we discussed in the previous section on teacher incentives). In a health RBF intervention in Burkina Faso, researchers found that the designers failed to target the incentives to certain groups of medical support personnel and health management committees even though they were working closely with the health workers who were being incentivized, which contributed to those actors' perceptions that RBF was just another form of regular development aid rather than something that could generate more systemic results.[160]

There is no conclusive evidence that proves that incentives work best at one particular level of the education system. However, based on operational experience, it seems to be more important to think about whether the incentives given to the different actors involved in the delivery of the service or programs are aligned rather than on who receives explicit incentives.[161] This requires designers to have a clear theory of change in mind, as well as to be aware of the political economy of the sector and of funds flow issues as, in some instances, there may not be a practical way to transfer funds to certain levels. Ultimately, the best incentive scheme for service providers, meso-level (district/province) stakeholders, or national-level stakeholders will not work if other agents' interests are not aligned with that incentive, as was seen in the Burkina Faso.

### Selecting and Pricing Indicators

In RBF, just as with traditional development financing projects, it is not always easy to know which indicators are the "right" ones, but indicators in RBF projects carry more weight because achieving them prompts the disbursement of funds. RBF indicators must strike a balance between cost, effort, feasibility, and ambition.[162] For more insight into the selection and pricing of indicators, we take a detailed look at designing disbursement-linked indicators (DLIs), which are the indicators that must be achieved for funds to be disbursed in RBF projects.

***DLI Analysis: The Basics***

We started by analyzing the disbursement-linked indicators (DLIs) that have been used in World Bank education projects. We classified indicators into four types (input, process, intermediate, outcome) to ensure more differentiation during the analysis. In Table 7, there are examples of each of these categories. For the purposes of this report, the primary difference between a process indicator and an intermediate indicator is that a process indicator generally reflects that an action or policy has taken place, but nothing additional has happened.

We focused on the World Bank's portfolio given the Bank's large share of RBF projects in the education sector, as well as the fact that it often acts as the implementing agency for other donors. The analysis covers 352 DLIs from 51 projects (investment project financing using DLIs and PforRs) from 2008 through June 2018. This is around 6 percent of the Bank's

160 Ridde et al (2018)

161 As mentioned, for example, in Olander and Högberg (2016)

162 World Bank (2017)

total education portfolio, which included 843 projects during that time period. Almost 94 percent of the funding in these projects was results-based, even when the financing instrument was not a PforR. We categorized the DLIs by topic (Table 8) and by their position in the results chain (Table 9).

*Table 8: Common Disbursement-linked Indicator Topics*

| TOPIC | EXAMPLE |
|---|---|
| **Inputs** | Budget, infrastructure, textbooks |
| **Data/Systems** | EMIS, school census data, annual reports |
| **Assessment** | Administering exams, examination commissions |
| **Teachers and Teacher Training** | Teacher management, teacher accountability, teacher deployment, teacher training and evaluation |
| **Enrollment, Completion, Retention** | Number of students, increase in students, number of students completing grade or training |
| **Quality Assurance** | Increased capacity, accreditation, performance benchmarks, readiness criteria |
| **School-based Management** | School grants, school management committees |
| **Curriculum** | Curriculum standards, curriculum framework |
| **Learning Outcomes** | Test scores, employment |
| **Policies/Frameworks** | Reforms, council or agency established/operational |
| **Other** | Scholarships, industry relationships, skills training |

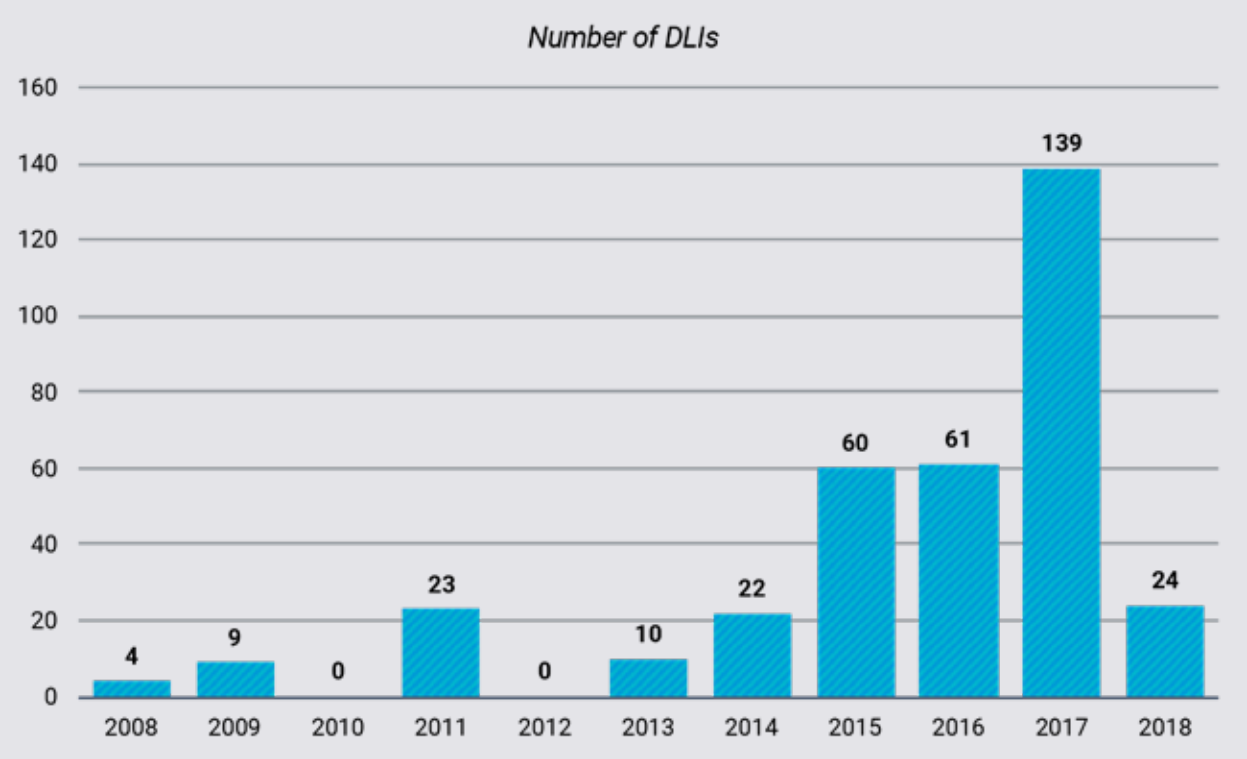
*Table 9: Examples of Indicators*

| POSITION IN RESULTS CHAIN | EXAMPLE |
|---|---|
| **Input** | Textbooks have been procured and delivered to targeted schools. |
| **Process** | An effective and relevant curriculum is in place. |
| **Intermediate** | The required number of pilot school inspections has been completed with reports published on the Ministry of Education's website. |
| **Outcome** | The recipient has demonstrated an improvement in student learning outcomes. |

Before delving into other questions, here are some basic details on the DLIs analyzed. The average number of DLIs per project was 6.9, with the median number being 6. However, there was some variation, with the bottom quartile of projects having 5 or less DLIs while the upper quartile had 9 or more. The values of DLIs are also quite variable. The mean DLI was worth around $18.8 million. However, the bottom quartile of DLIs cost $12 million or less each and the top quartile $29.5 million or more each. Some of the DLIs were worth huge sums of money, especially in P4R projects. The highest-valued DLI was worth US$ 341.5 million and belonged to the Nigeria Basic Education Project.

Regarding time trends, it is too early to tell, since our sample is limited to 51 projects. The only result worth highlighting is the growing popularity of results-based financing, as evidenced by the increase in the number of DLIs over the past few years.

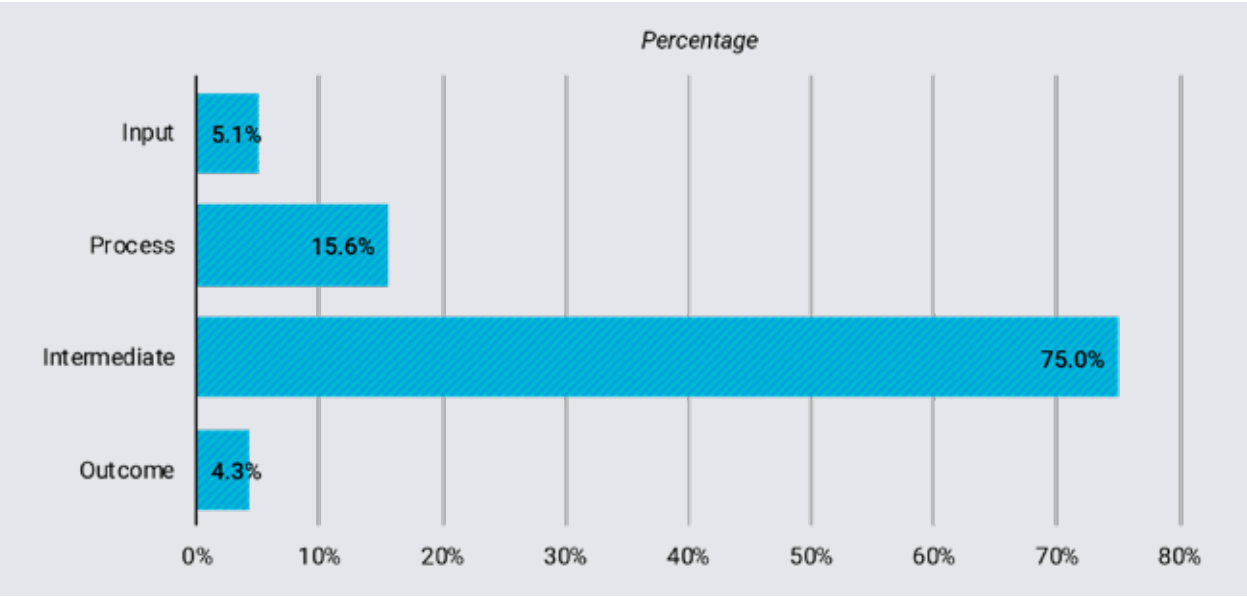
*Figure 7: Number of DLIs by Year*




***DLIs and Results Chains: Few DLIs Focus on Outcomes***

Our survey results showed that most practitioners believe that incentivizing a mix of results is the most effective design, and our DLI analysis results supports this. According to the DLI analysis, most indicators are set at the intermediate or process level, with very few at the input or outcome level. Overall, as seen in Figure 8, the majority (75 percent) of DLIs are focused on intermediate outcomes, meaning that they require an improvement or strengthening of something, such as teacher training. Very few DLIs focus on inputs (5 percent), while even fewer are set at the outcome level (4 percent).

*Figure 8: DLIs by Position in Results Chain in World Bank Education Projects*




While the survey respondents indicated that inputs were worth financing (see Figure 10), our DLI analysis revealed that inputs were very rarely used as DLIs in World Bank education projects. Even indicators related to traditional inputs such as textbooks tended to focus on quality aspects, such as timelier delivery of textbooks. Similarly, DLIs related to school construction were generally more about ensuring quality improvements than about building more schools.

In a recent World Bank evaluation of the PforR instrument (the most widely used financing instrument that directly ties financing to indicators within the World Bank), roughly 48 percent of disbursement-linked indicators across all sectors were defined as results such as capacity building and institutional development, which do not qualify as final outcomes.[163]

This is most likely due to pragmatism on the part of project designers. In fact, the ADB recommends that teams focus on institutional strengthening when selecting DLIs to capitalize on the potential of RBF to strengthen systems and institutions.[164] In the responses to our survey, there were mixed opinions, with a few respondents echoing, "RBF is too often used for processes and inputs; because these are seen as 'easy' and so can provide a flow of funds. But fund flows should be smoothed with other mechanisms. Processes can be important if they really represent a change in the way a system is operating." However, the majority of survey takers indicated that context was incredibly important and that "good projects finance different steps in the result chain, not only final outcomes." The importance of incentivizing throughout the results chain will be illustrated in later examples.

In contrast, some studies have argued that outcome-based indicators are associated with better results since they are agnostic on the activities required to produce those outcomes.[165] Arguably, this allows agents to find the best way to achieve the outcomes. However, as previously discussed, outcome indicators are also those over which

163 IEG (2016)

164 ADB (2016)

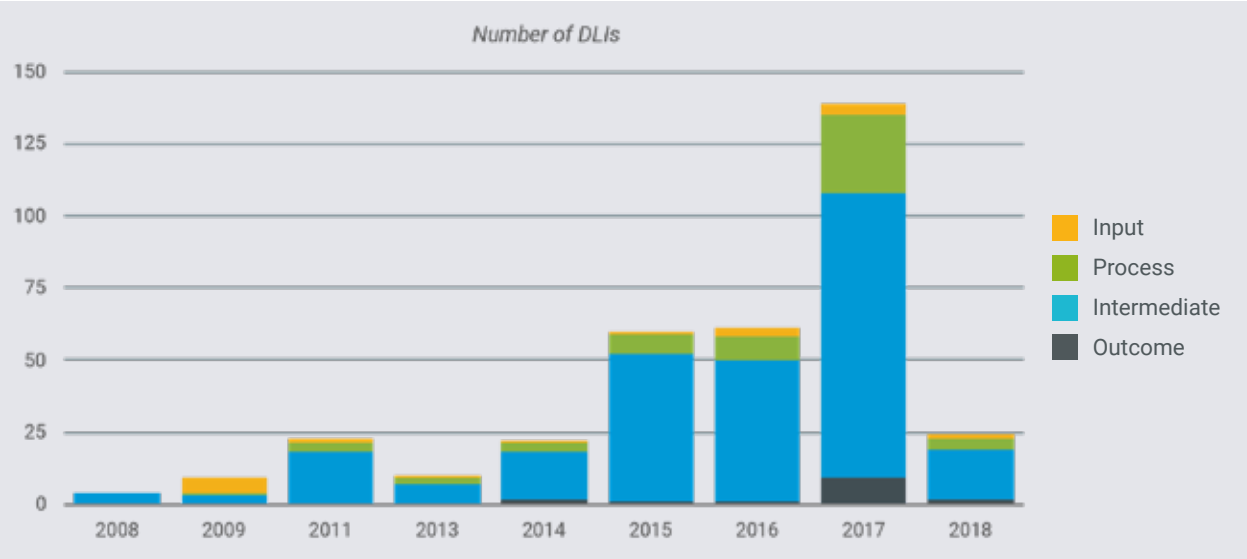
165 Holzapfel and Janus (2015)

agents have the least control. In education interventions, an example of an outcome indicator is literacy or student learning as measured by test scores. There are many factors that can affect these indicators, thus undermining the link between the effort of the implementing agent and the results that it has to show for it. As mentioned in the previous section on teacher incentives, test scores are prone to random variation due to shocks, cohort characteristics, or school size.[166] This is why many survey respondents believed that it is crucial to extend RBF further up the results chain, such as intermediate outcomes and even process indicators and inputs.

A good way forward may be to focus not just on the position of an indicator in the results chain but also on how much control the agent has over achieving it. This is a principle that can also apply to the rest of the results chain since there are also other types of indicators that are more controllable than others. Therefore, it is important to think seriously about what determines changes in any indicator year to year. For example, enrollment rates are a common indicator in many education projects, but it is contingent both on the number of students enrolled and on how the denominator (the universe of school-age students) is measured.

Another lesson that we learned from our DLI analysis included the importance of valuing indicators based on their leverage (or ability to unlock processes and progress) rather than on their value for money. It is worth mentioning that DLIs are often tied to larger systemic change. For example, one objective of an ongoing basic education project in the Dominican Republic is to increase the country's capacity to recruit and train primary and secondary school teachers. Attached to this objective are a series of specific DLIs, such as the development and dissemination of professional standards for secondary school teachers.

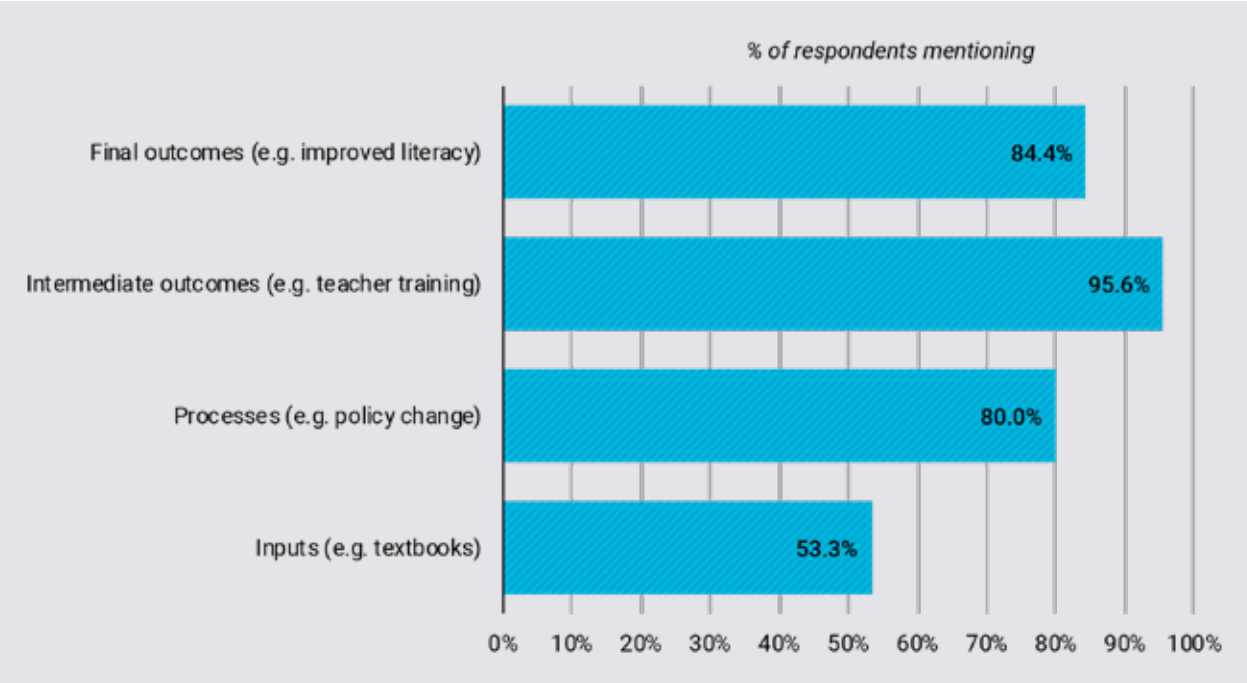
***Figure 9: DLIs by Position in World Bank Education Projects***




*Source:* Authors' analysis of DLIs in World Bank education sector projects

166 Murnane and Ganimian (2014)

*Figure 10: Types of DLIs Worth Incentivizing*




*Source:* Authors' survey of opinions on RBF of development agency staff working in education

Given how difficult it can be to identify a clear formula that leads to strong learning outcomes, the best DLIs to select often depends on understanding the results chain. As one survey respondent wrote, "If a major constraint to equity in a particular country is the near-permanent delay in the production and delivery of mother-tongue-instructional materials for ethnic communities, tying disbursement to this input could be powerful. New or revised processes can be sensitive and challenging to design or implement so tying disbursement to process indicators could be key in a situation like this." This also speaks to the idea that deciding where an indicator fits in the results chain is hard because it depends on perspective—an input in one project may very well be an output in another.[167]

***Examples of Results Chains: Bangladesh, Lebanon, and Tanzania***

Ultimately, in RBF projects, the results chain is of paramount importance, and it would be helpful for teams to ensure that they have all the necessary context-relevant information about the results framework.[168] Here we present some examples of results chains and DLI progression in World Bank RBF projects in Tanzania, Lebanon, and Bangladesh. They illustrate how and where disbursement-linked indicators can be inserted along the chain on the assumption that incentives can potentially unlock identified bottlenecks.

167 ADB (2017), Clist and Verschoor (2014), and Holzapfel and Janus (2015)

168 Sabarwal et al (2016)

*Figure 11: Results Chain Example from Tanzania*

**Results Chain Example: Tanzania**

| LEVER | INPUT/ACTIVITY | INTERMEDIATE OUTCOME | FINAL OUTCOME | HIGHER ORDER |
|---|---|---|---|---|
| **Strengthen Performance—Transparency** | Official School Ranking | School ranking released | Identification of lagging schools, students, and teachers [for better planning and focused attention] | Improved student learning outcomes in PSLE & CSEE examinations |
| | National 3R Assessment | No. of schools participating in the 3R assessment | | |
| **Motivate through incentives** | School Incentive Grants (SIG) – performance based | No. of schools receiving performance-based incentive rewards (DLI) | Increased teacher effort measured through classroom presence (PDO) | |
| | Non-financial performance incentives for teachers | Teacher awards announced yearly to high-performing teachers | | |
| **Improve Teacher Conditions** | Clear backlog of claims | No. of outstanding Teacher claims older than three months | | |
| **Provide Support where required** | School improvement toolkit | No. of schools that receive the toolkit | | |
| | 3R teacher training program | No. of teachers trained | Improved teacher proficiency 3R subjects (PDO) | |
| | Student-Teacher Enrichment Program (STEP) | No. of schools participating in STEP | Improved student performance in 3R assessment (PDO) + (DLI) | |
| | Timely Delivery of Adequate Capitation Grants | Percentage of schools receiving capitation grants on time (DLI) | Improved textbook student ratio | |

*Source:* Big Results Now in Education Program, Project Appraisal Document, World Bank 2014

*Figure 12: Results Chain Example from Lebanon*

**Results Chain Example: Lebanon**

| LEVER | INPUT/ACTIVITY | INTERMEDIATE OUTCOME | FINAL OUTCOME | HIGHER ORDER |
|---|---|---|---|---|
| **Equitable Access** | School construction and rehabilitation | No. of school-aged children (3–18) enrolled in formal education (DLI) | Increase in the proportion of school aged Lebanese and non-Lebanese children (3-18) enrolled in formal education (PDO) | Improved access, quality and a stronger education system to respond to the refugee crisis |
| | Psychosocial program for students to help with re-integration into formal education | No. of children and youth whose registration fees for public formal education and ALP are partially or fully subsidized | | |
| | | No. of public schools newly built or expanded to meet quality standards specified in GoL's Decree 9091 | | |
| **Enhanced Quality** | Training for educators | % of children and youth aged 3–15 above the corresponding graduation age who have completed a Cycle | Increase in the proportion of students passing their grades, and transitioning to the next grade (PDO) + (DLI) | |
| | School grants | | | |
| | M&E of teacher quality, learning outcomes, and learning environments | Proportion of students transitioning grades (DLI) | | |
| **Strengthened Systems** | EMIS | Unified framework for data management, data collection protocols, and compliance systems endorsed and operational (DLI) | Timely and robust data available for evidence informed policy-making and planning (PDO) | |
| | Capacity building | CERD adequately capacitated and equipped to develop interactive content and e-platform | | |
| | Revised curricula | Curriculum revised to improve quality of learning | | |

*Source:* Adapted from Reaching All Children with Education in Lebanon Support Project, Project Appraisal Document, World Bank 2016

In Tanzania, the government had already participated in an intensive retreat to identify the primary results that it wanted to achieve as part of its education reform initiative. Thus, the team focused on designing DLIs that would: (i) link incentives to key points within the program results chain; (ii) focus incentives as closely as possible on the key actors accountable for their attainment; (iii) be simple and manageable in terms of their number and framing; and (iv) have a high likelihood of being achieved within the specified timeframe and within the control of the government.[169]

In Lebanon, the team sat down with their government counterparts to discuss the key areas in the education system that required immediate improvement. These areas became the key levers, or pillars, of the project. As a result of these discussions, the DLIs were structured to follow a logical progression down the results chain with specific reference to the Syrian refugee crisis, because of which many Syrian children were out of the formal education system. For example, the first DLI was the incorporation of all children into the education system. The second was for all of these children to complete the school year and transition through into the next year throughout the grades. The remaining DLIs bolstered other key pillars in recognition that enrollment, completion, and retention are not sufficient achievements on their own.[170]

In the Bangladesh Primary Education Project, DLIs were introduced during the third iteration of the project. In this case, the government already knew how to manage donor funding, and many of the reforms under the second project were being rolled into the third.[171] The first DLI was the introduction of a five-year action plan to improve the Grade 5 completion exam, and subsequent DLIs focused on revising the exam and piloting it with incremental increases in the number of competency-based exam items. Ultimately, the goal was not only to improve the exam but to also ensure that the results were analyzed and disseminated in a timely manner.

169 World Bank (2014b)

170 World Bank (2016)

171 World Bank (2011)

*Table 10: Sample DLIs from the Bangladesh Third Primary Education Development Project*

| DLI | BASELINE | YEAR 0 (May–June 2011) | YEAR 1 (April/May 2012) | YEAR 2 (April/May 2013) | YEAR 3 (April/May 2014) |
|---|---|---|---|---|---|
| **3. Grade 5 Completion Exam:**<br>Improving the quality of primary completion exam and the regular measurement of learning | Grade 5 completion exam implemented for all primary school students in 2009.<br>Content focused on testing students' memory more than ability to use subject knowledge | A 5 -year Action plan for improvements in Grade 5 Completion<br>Exam developed by NAPE and approved by MOPME and including revising test items to gradually transform exam into competency-based test<br>New test items developed by NAPE on selected competencies and piloted with accompanying guidelines for pilot test administration and training of test administrators | Revised 2011 Grade 5 Completion Exam, based on action plan and pilot results implemented, incl. guidelines developed for markers and training of markers<br>**Analysis of 2011** Grade 5 completion exam results and content completed by DPE and NAPE and results disseminated | Action plan implemented with at least 10% of items competency-based introduced in the 2012 Grade 5 exam and an additional 15% of competency-based items piloted<br>**Analysis of 2012** Grade 5 completion exam results and content completed by DPE and NAPE and results disseminated | Action plan implemented with at least 25% of items competency-based introduced in the 2013 grade 5 exam and an additional 25% of competency-based items piloted<br>**Analysis of 2013** Grade 5 completion exam results and content completed by DPE and NAPE and results disseminated |

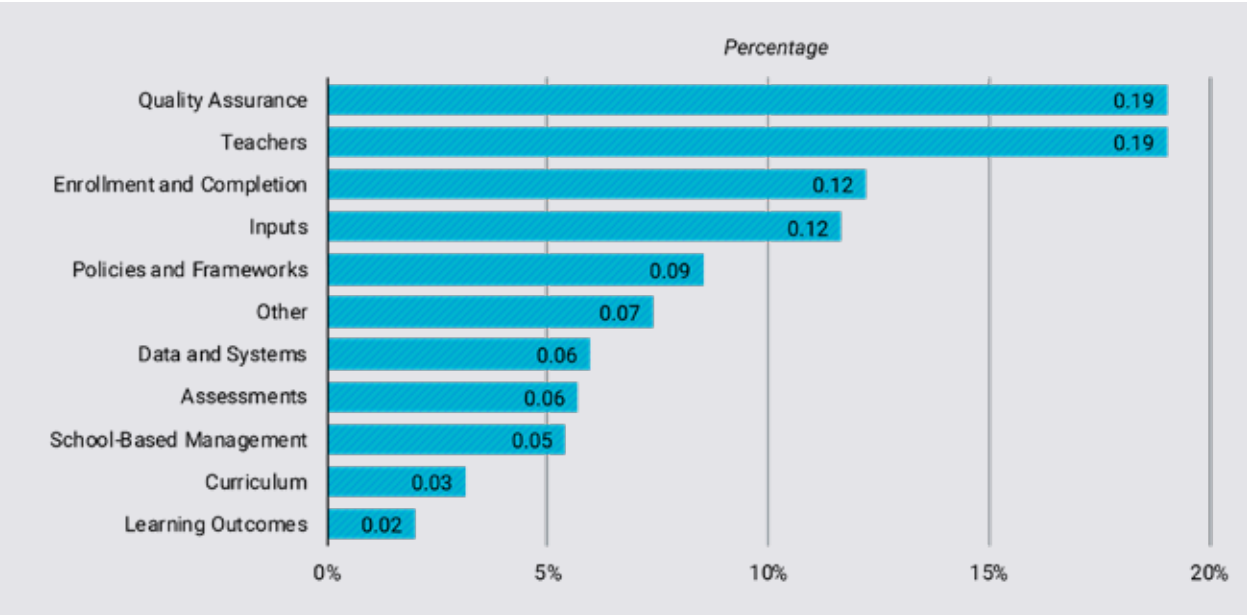
| PROTOCOL |
|---|
| **Definition: The Grade 5 Action Plan specifies the number of new competency-based items to be introduced each year, with the aim of achieving a fully competency-based exam by end- 2016.**<br>Analysis of results includes:<br>(i) analysis of pass rates by gender, subjects, Upazilas conducted by DPE; and<br>(ii) analysis of NAPE of marking and scoring of a sample of answered scripts in selected Upazilas.<br>Source: Action plan as approved by DG, NAPE and MOPME; sample of test items and questionnaire of grade 5 exam; test analysis reports by DPE and NAPE. |

*Source:* World Bank (2011b)

All three of these examples illustrate the theories of change within the projects and how incentives were used to promote the achievement of higher order goals.

In our DLI analysis, we also looked at the thematic focus of DLIs to see if there were any patterns in the types of DLIs that teams preferred. While the DLIs covered a range of topics, the most common ones were: (i) teachers and teacher training; (ii) quality assurance; (iii) enrollment, retention, and completion rates; (iv) inputs such as textbooks or budgets; and (v) policies/frameworks. There was nothing to suggest that these were unusual, and it is possible that non-RBF projects might also favor the same types of indicators.

*Figure 13: DLIs by Topic*




As previously mentioned, the types of achievable results will heavily rely on the country context and, in particular, how much control an agent has over the result. Although there is no clear way to select specific DLIs that will guarantee transformation, other researchers and agencies have put together some criteria to help practitioners to select DLIs and to assess the quality of those indicators (see Table 11 and Table 12).

*Table 11: SIDA's Checklist for Choosing Indicators in RBF Indicators relate to the desired outcome*

- ☐ Indicators are neither too many nor too few
- ☐ Indicators are based on what is already there and needed anyway
- ☐ Indicators are precisely defined with clear protocols
- ☐ The measurability and periodicity of the indicators are fully verified
- ☐ Baselines are determined and verified
- ☐ Any targets are sensibly set
- ☐ Timeliness of data informs planning, budget, and disbursement schedules
- ☐ Flexibility is built into the agreement

*Source:* SIDA's internal guidelines for RBF

*Table 12: Criteria for Assessing the Quality of Disbursement-Linked Indicators*

| CRITERION | KEY QUESTION | CONSIDERATIONS FOR RESULTS-BASED APPROACHES |
|---|---|---|
| **1) Focus on results** | Do indicators ensure a focus on results? | • The indicators can measure results (outputs and outcomes) or processes (inputs and activities) |
| **2) Control** | Can results be influenced by and plausibly associated with the intervention? | • The extent to which incentivised actors have control over achieving the intended results<br>• The extent to which results can be attributed to the intervention<br>• The institutional setting of incentivised actors |
| **3) Financial incentives** | Can intended effects be maximised? | • The extent to which financial amounts reflect 'value for money', policy leverage, risk or other considerations<br>• Whether disbursement is scaled in proportion to performance or conditional on achieving a threshold level |
| **4) Measurability and verifiability** | Are indicators reliable, consistent over time and independently verified? | • The relationship between the indicator and the underlying objective of the programme<br>• The data quality and source (administrative data or survey data)<br>• The way verification is organised (independent or not) |
| **5) Unintended consequences** | Can unintended effects be minimised? | • The extent to which indicators allow gaming (active manipulation of the indicators)<br>• The extent to which indicators lead to distortions (indirect consequences of overemphasising or neglecting policy choices) |

*Source:* Holzapfel and Janus 2015

***DLIs and Short Project Timelines***

Another important consideration when selecting indicators is the timeframe within which indicators are expected to be achieved. Nowadays, many development projects are scheduled to last only four to five years (for example, according to the GPE guidelines, grants with the variable part are to last three to four years), and teams must be realistic about the type of result that can be achieved within that timeframe. If the project is to be implemented in a country that is struggling to meet its basic education goals, then setting outcome targets may only be setting things up for failure. The results of our survey and interviews also point to this. One respondent wrote, "In a recent grant to Cambodia … the implementation timeline was too short for RBF to be used meaningfully." Both DFID-funded evaluations of projects in Rwanda and Ethiopia noted that the projects' life cycles were probably too short to generate any long-term changes. One possible solution might be to apply RBF either after traditional financing projects, as in the Bangladesh example, or as a way to incentivize results further down

the chain directly after the completion of an existing RBF project.

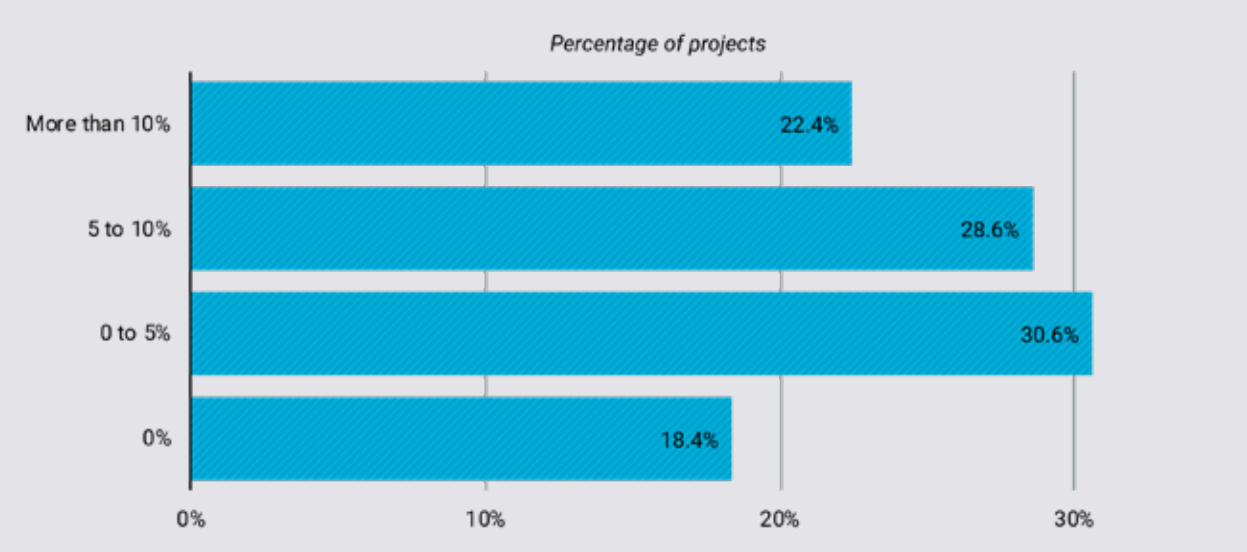
***Pricing DLIs: Three Hypotheses and a Heuristic***

There does not seem to be any consensus on how to price DLIs, though researchers have suggested several criteria that could be used, including (i) value for money; (ii) leverage effects; and (iii) additional risks for partners.[172]

Using the value for money criterion involves pricing a DLI in proportion to the benefit or "value" of the activities required to attain it. As discussed above, using the leverage criterion relates to whether a DLI can unlock results further down the results chain or is complementary to other DLIs. Using the criterion of spreading out risks involves distributing the disbursements amongst several DLIs to avoid an all-or-nothing situation. For example, a project could combine some DLIs that are easier to attain with some that are harder.

There is some support for these hypotheses. Based on our interviews with practitioners and our DLI portfolio analysis, we found the most common practice for pricing a DLI in World Bank RBF projects was to use a simple heuristic: divide the number of DLIs by the total amount set aside for DLI disbursements. Thus, if a project had 10 DLIs and the financing allotted to DLIs was US$100 million, then teams would roughly price each DLI at US$10 million. The table below shows the average dispersion of DLI indicators, measured in percentage of the total funds. In around 18 percent of projects, all of the DLIs were priced equally, and therefore the version is 0 percent.[173] Around 50 percent of projects have an average dispersion of 5 percent or less.

***Figure 14: DLI Value Dispersion by Project***




172 Holzapfel and Janus (2015)

173 In other words, the standard deviation across DLIs within the project was 0.

In general, we found that most of the variation in DLI pricing came from differences between projects rather than within projects. In other words, different DLIs within the same project tended to be closer in price than similar DLIs in different projects. This approach suggests that teams are indeed trying to spread out the risk and minimize the potential negative impact of countries failing to achieve the DLIs by ensuring that all DLIs are worth roughly the same in terms of disbursements.

Another pricing mechanism that came up during our analysis, survey, and interviews was to "make the most important thing the most expensive." This adds some evidence to the "value for money" and "leverage effects" hypotheses. In RBF, the actual expenditures of an activity can be delinked from disbursements and the activity can then be priced for its perceived worth. While a policy reform may cost nothing, it may be a major improvement in terms of enabling results further down the results chain and, thus, could be priced accordingly.

Overall, this evidence suggests that project teams do not price DLIs in proportion to their real cost since, otherwise, there would be more variation within projects and less across similar projects. Instead, they seem to focus on how to spread out disbursement risks and on the perceived value of DLIs as a way to produce results further down the results chain.

As we have shown, DLI pricing in World Bank RBF education projects does not follow any particular formula nor is it clear how much a DLI should cost to create "sufficient" incentive. In the Sri Lanka Education Sector Development Program (see Table 13), which is set to close in 2019, there are nine disbursement-linked results areas (DLRs), which cover a total of 37 indicators. The financing for each DLR is either US$10 million or US$32 million, but no explicitly articulated rationale exists for these allocations. Presumably, if the donor wanted to signal the importance of one indicator over another, there would be more variation between the allocations. Instead, disbursements are relatively evenly spaced out per year, with lower disbursements in the final two years. This makes sense given that most of the legwork to achieve final indicators would have been done in the previous years.

In addition, the program budget shows a total of US$200 million allocated to DLIs, which is a sizable amount and is probably enough to incentivize action. However, in an early childhood development project in Jamaica, each DLI was only worth US$180,000, yet the government was still incentivized to achieve the targets.[174]

174 World Bank (2008)

*Table 13: Disbursement Table of DLIs for Sri Lanka Education Sector Development Program*

| # | DLR | BANK FINANCING ALLOCATED TO DLR ($US M) | YEAR 1 (US$ M) | YEAR 2 (US$ M) | YEAR 3 (US$ M) | YEAR 4 (US$ M) | YEAR 5 (US$ M) |
|---|---|---|---|---|---|---|---|
| 1 | **Pass rates for GCE O level examinations increased** | **10** | | | | 5 | 5 |
| 2 | **Pass rates for GCE A level examinations increased** | **10** | | | | 5 | 5 |
| 3 | **Pathways from school to TVET developed-Technology Stream commenced and implemented at GCE A levels.** | **32** | 8 | 8 | 8 | 4 | 4 |
| 4 | **Secondary schools upgraded to offer all subject streams** | **32** | 8 | 8 | 8 | 4 | 4 |
| 5 | **Enrollment in GCE A levels Science Stream increased** | **10** | | | | 5 | 5 |
| 6 | **Enrollment in GCE A levels Commerce stream increased** | **10** | | | | 5 | 5 |
| 7 | **Principals and deputy principals trained** | **32** | 8 | 8 | 8 | 4 | 4 |
| 8 | **Institutional capacity at MOE and provincial levels strengthened.** | **32** | 8 | 8 | 8 | 4 | 4 |
| 9 | **Improved transparent and efficient procurement** | **32** | 8 | 8 | 8 | 4 | 4 |
| **TOTAL** | | **200** | 40 | 40 | 40 | 40 | 40 |

A/L = advance level, DLR = disbursement-linked results, GCE = General Certificate of Education, MOE = Ministry of Education, TVET = Technical and Vocational Education and Training.

*Source:* ADB (2013)

### Box 2: Overpriced Indicators in Uganda

In Uganda, a health intervention led by the Dutch NGO Cordaid used RBF to incentivize the use of maternal and neonatal care services. The program faced budget constraints after only two years of implementation. As a result, the most expensive subsidy for outpatient consultation services was cut in half. Initially the facilities complained about the reduction, but attendance rates did not go down. Cordaid originally hypothesized that facilities would simply start charging higher user fees to cover the gap, but upon further investigation, it was discovered that the initial incentive had dramatically improved the quality of public facilities so that they could now compete with private not-for-profit health centers. This improvement in quality was able to generate some price competition. Perhaps the most interesting conclusions from this example is that it showed Cordaid that the outpatient consultation services indicator had likely been originally "overpriced" and that indicator pricing, in general, warrants more thought.

*Source:* RBF Health (2017)

In general, there are no distinct trends for pricing DLIs based on themes. Our analysis showed the average price of a DLI based on its position in the results chain. As can be seen in Table 14 below, process DLIs have a lower value than intermediate outcomes, and intermediate outcomes have a lower value than final outcomes. Intuitively it makes sense that final outcomes would be the costliest indicators since they are the most difficult to achieve, but it is unclear why process indicators are, on average, worth less than input-related ones. More analysis of individual DLIs in the process category is needed to understand how teams are pricing them in relation to input-related DLIs, although, based on our overall DLI analysis, there may be no distinguishable methodology being applied.

***Table 14: Cost of Indicators Based on Position in the Results Chain***

| POSITION | AVERAGE COST (US$ M) |
|---|---|
| Process | 12.7 |
| Input | 26.3 |
| Intermediate | 18.5 |
| Outcome | 39.5 |

While these examples give an idea of how teams have priced DLIs, they are by no means based on any science. In reality, project teams often make a judgement when it comes to costing, which is usually done jointly with the government and other relevant development stakeholders. To provide teams with more guidance, additional analysis could be conducted to identify which types of indicators disburse more easily, based on their theme and position in the results chain.

***DLIs: Scalability and Disbursement Models***

One of the advantages of RBF as a funding mechanism is its flexibility. Making DLI disbursements scalable can be a useful way to mitigate the risks faced by the borrowing governments, especially those facing liquidity

constraints. Many projects have scalable DLIs, meaning that even if the borrower only partially achieves the DLI, it can request the disbursement of an agreed proportion of the total value. In some projects, DLIs are also not time-bound. We found that a little over half of the DLIs in our DLI analysis were scalable (55 percent). For roughly 15 percent, it was unclear whether the DLI was scalable or not. In Table 15 below, there are examples of the most commonly used disbursement models in RBF education projects.

*Table 15: Advantages and Disadvantages of Disbursement Models*

| DISBURSEMENT TYPE | FEATURE | ADVANTAGE | DISADVANTAGE |
|---|---|---|---|
| **Target based** | Allocating a fixed amount per DLI and disbursing once annual targets are met. | Easily understood, can be scalable. | Difficult to set because it is hard to know how much progress will be made on a yearly basis.<br>Often ends up penalizing borrower for lack of progress instead of rewarding the country for making good progress. |
| **Baseline based** | Rewarding progress over the current baseline (can be a rolling baseline). | Easily understood, can raise the bar every year as long as there is progress. | Provides same incentive regardless of degree of progress, e.g. any increase over the baseline will trigger full disbursement. |
| **Progress by unit** | Rewards progress proportionally on basis of agreed price/ reward per unit | Most flexible approach, most common method used in RBF projects. | Pricing can be difficult to ascertain, may need to have a cap on maximum amount that can be disbursed. |

*Source:* SIDA's internal guidelines

***Zero/Global DLIs***

Previously, "zero DLIs" were described as an option to help governments (and other entities) to manage liquidity risks by creating indicators that are achievable and thus can bring in financing that can fund efforts to achieve other, more difficult indicators. While not formally part of any guidance or research, this is a common practice used by project teams to introduce RBF into contexts where it has never been used before in an effort to acclimate them to it.

Another practice often adopted by project teams consists of "global DLIs," which are DLIs that must be met for the project to continue. In other words, if they are not achieved, then nothing else can happen. While liquidity constraints and the scalability of DLIs are important to consider, in some projects these are insufficient criteria to ensure meaningful progress. In short, flexibility can only overcome so much.

The Sindh Education Project included a global DLI that was conditional on the implementation of merit-based recruitment of public school teachers. This became a necessary condition for all disbursements, in other words, there would be no disbursements if the DLI was not met. The rationale for that particular approach was that the DLI was considered by both the Government of Sindh and the World Bank as central to increasing education access, improving education quality, and improving sector governance and as an indicator whose achievement could be strongly assisted by Bank financing. This particular DLI also entailed a high implementation risk and thus was worth turning into a global DLI.[175]

In conclusion, as shown, there are many ways to select and price indicators. They can be negotiated between the donor and the government/recipient, they can be set by international benchmarking, they can be aligned with national objectives (as part of a development plan, for instance), they can fit into a cost-effectiveness framework, or they can even be decided by a panel of experts.[176] It is clear that there is no consensus on which method is most effective. Thus, more research into optimal indicator selection and pricing practices would be useful to guide practitioners on the types of indicators that are likely to lead to the best outcomes and how much those are worth.

## Adaptive Implementation

Implementation is a lesser understood and researched part of RBF projects. This is probably because most RBF education projects are still being implemented so it is not yet possible to analyze what lessons can be learned. In addition, the nature of implementation is very context-specific and requires more robust qualitative research.

The main point of this section is to highlight the need for adaptive and flexible implementation. In RBF, the way in which an indicator is achieved is very open and can require careful management and supervision. While this is where innovation and autonomy on the client side can occur, in practice, this is usually not the case. RBF projects in education often require technical assistance from donors for the desired results to be achieved.

### Monitoring and Information Systems

Monitoring and information systems are critical for results-based financing, which is based on the ability to accurately monitor and verify indicators. Thus, these systems must be in place before countries can put RBF into practice. An education management information system (EMIS) can serve multiple purposes, but the primary goal of teams is to use EMIS data as a way to assess the strengths and weaknesses of the education system.

In most instances, monitoring and information systems are needed to ensure the disbursement of funds, but RBF can also be used to establish or improve existing monitoring and information systems.

***Purpose of Education Management Information Systems (EMIS)***

Given that RBF requires strong monitoring and information systems to ensure that indicators are accurately tracked, EMIS can be used as a disciplinary tool, in other words, if governments do not meet agreed targets, then they face punitive measures and donor funds are not disbursed. Conversely, governments face positive measures when indicators are met. This is a somewhat overly dichotomous depiction of the process, but it illustrates one of the fundamental concepts behind RBF, that the threat of not getting the funds encourages governments to make all possible efforts to achieve indicator targets.

175 IEG (2013)

176 Cruz-Aguayo and Martínez (2016)

In reality, most practitioners think that the best way to use an EMIS in an RBF project is to see it as a feedback mechanism. Rather than using a punitive lens with regard to the non-achievement of targets, most of the practitioners that we interviewed thought that it was more important to know why the targets were not being met. An interviewee from the ADB said that the "EMIS and the verification process should always be used to strengthen the system."

For example, in a health project in India funded by the ADB, the team chose two outcome-level indicators, one which was chosen even though data from the country's systems were not reliable and there were no funds available to conduct a survey. There was a regular household survey, but it was fielded only every three to four years. Given these constraints, the team proposed that the EMIS should be based at the health center level to give the centers a strong incentive to collect and report data. While these data might not be 100 percent accurate, the ADB team wanted to see a general upward trend rather than exact percentile increases. While this may not have been a very scientific approach, it was practical, and it gave the government a strong motivation to strengthen the EMIS, which was a positive result in and of itself, as well as a general improvement in health services.

***Building Systems for RBF***

The higher a country's income level, the more likely it is to have EMIS in place. However, some low-income countries are also able to develop an effective EMIS. REACH recently awarded a grant to Haiti to enable the development of an EMIS and, therefore, create the pre-conditions for an RBF project in the future. Rather than start with an RBF intervention and scramble to create information and monitoring systems from scratch, the intervention started by engaging in a dialogue with the government on the following question: what results did they want to see in the education system, and what information and monitoring systems would be required?

Based on this discussion with government counterparts and education sector stakeholders, the program put together a results framework that led to the development of a quality assurance system (QAS). This is now guiding the government's efforts to strengthen the national statistics agency and the EMIS.[177] This is a good example of how RBF can be framed when information systems and capacity are lacking. It is not always about jumping to the finish line and financing outcomes but rather about thinking about what needs to be in place before an RBF intervention can happen.[178]

***Level of Complexity Needed for Systems***

Another design issue is the level of complexity required of an information and monitoring system. There are many examples in the development literature of complex, expensive information systems that were funded and initiated by donors but were not successful.[179] A classic example is SISTAFE — the PFM system that was developed in Mozambique. Despite using state-of-the-art software and receiving millions in donor funding, it did not take off.[180] It was slow to be adopted by line ministries and, to date, is still not universally used.

There are many reasons for this sort of dynamic. First, sometimes technology is prioritized over purpose. IT systems and software are much easier to develop and replace than organizational culture and practices, so it can be tempting to build an EMIS from scratch but then to find that no one uses it. Second, the donor's needs can sometime be

177 Barón and Adelman (2018)

178 World Bank (2017)

179 Andrews et al (2017)

180 Andrews (2013)

prioritized over those of the country. Because information systems are crucial for the effective implementation of RBF, there is the danger that the donor will invest in designing a country's EMIS with the sole purpose of serving a program or intervention. This just ensures that it will never be used again once the program is over.

In some contexts, the challenge is to build a system from scratch because nothing else exists. In these contexts, it is worth it to build only simple and easy to maintain systems. For example, the REACH-funded Haiti intervention aimed to develop a functional EMIS to solve the country's problems, not just to serve donors or programs. In Niger, the Ministry of Education has limited data sources and relies mainly on school censuses and statistical yearbooks,[181] and likely would not require a more complex, high-tech system at present.

However, if the country in question has plenty of reliable data, strong capacity to interpret data, and a functioning EMIS, then the challenges are different. For example, REACH recently funded a grant to fund an intervention in Colombia to redesign the country's EMIS around the objective of education quality, thus creating SIGCE (the Management System for Education Quality). The team avoided pitfalls such as creating a whole new system with the sole purpose of enforcing accountability since that would have imposed more administrative costs on different government agents with no clear benefit. Instead, the objective was to design a useful and user-friendly tool to enable stakeholders to understand the weaknesses and strengths of the education system and to guide policymaking.[182] Essentially, the new system was meant to be a management tool for education stakeholders to use proactively.

The Colombia program started by holding focus groups to discover the concerns and problems of agents at different government levels and to have them state their vision for the national information system. Only then did the team start work on the technical design of the indicators and software. Although the system is still being piloted, it is an example of how to undertake an EMIS reform by engaging all of the actors involved. Recent evidence suggest that this broad engagement is one of the determinants of whether a reform succeeds or not (rather than a lone political champion).[183]

***Monitoring Options***

There are various ways in which data can be collected and reported in order to monitor the achievement of indicators. Generally, if government information systems and administrative data are available, then this can be the cheapest option.[184] However, this can be problematic if the data are unreliable, inaccurate, or just not updated frequently enough. If the agency in charge of the data benefits from the RBF intervention, there is also a risk of gaming and cheating.

The alternative is to create a parallel structure to gather the information required to implement the RBF intervention. While this may be more reliable than working with existing government systems, it can be a missed opportunity to strengthen state capacity, and it goes against the aid effectiveness agenda.[185] Our interviews also revealed practitioners' frustration that, even though most education projects include an EMIS as a component, there is very rarely a functioning EMIS in-country after the projects have closed. One interviewee echoed the idea that a complicated and high-tech EMIS may be unnecessary and difficult to maintain, particularly at the school or health center level.

181 Majgaard et al (2018)

182 Cerdán-Infantes (2018)

183 Andrews et al (2017) and Andrews (2013)

184 Holzapfel and Janus (2015)

185 Holzapfel and Janus (2015)

Instead, it might be more beneficial to simply use the existing manual systems and improve on them by including more data fields to be collected or by strengthening the validity and reliability of the data collection process.

An interesting innovation in terms of data collecting and monitoring is the use of open source techniques. Open source data collection was pioneered by Cordaid in several RBF projects in Africa.[186] For example, in one of their health interventions, health centers are requested to draft their business plans and then enter them in the online platform that is open for everyone to access. Health centers then report their performance data in a monthly or quarterly fashion, and this is also published on the platform. Since the data are entered by the centers themselves in a decentralized manner (through computers, smartphones and tablets), this makes the process easier. Once the data are entered, they are validated externally, for example by the donor team. The local community can also verify in the system that the services are being provided. If the results add up, funds may then be disbursed.[187] It is important to note that the system still relies on monthly verification visits to ensure that the data are not being misreported.[188]

### Verification

Even with well-designed monitoring and information systems in place, RBF, by its very nature, depends on the verification of results. Without credible systems that can evaluate whether a target or result was attained (whether run by the government or otherwise), RBF will not work. One of the reasons that is often cited for the use of RBF is that it is a way to align the interests of the principal (donor) and the agent (recipient),[189] but if there are no verifiable data that both sides can agree on, this theoretical relationship is broken. The agent has no way to convey to the principal which results have been attained, and the principal has no way to know whether any of the information that it receives is reliable.

Once donors and country governments can agree on a system for collecting and reporting the data pertaining to the intervention, the most important question is who will be responsible for carrying out the verification. This can be done by central governments or line ministries, local service delivery agencies, or external firms or NGOs.[190] Whoever carries it out, the most common option is to use EMIS data, if available. This was the approach used by the Ethiopia intervention piloted by the DFID, which was discussed earlier.[191]

Using international organizations to carry out the verification is usually the costliest option since it is done through direct supervision. Conflict of interest can be a problem in these situations since the donor can sometimes be under pressure to disburse the funds because disbursements are often a metric of success. Our interviews with practitioners indicated that using this type of verification can send a signal to the government that the donor distrusts them, though sometimes, it is the government itself that requests that type of verification. In any case, using international organizations to carry out the verification is unsustainable over the long run because it does not build local capacity nor does it help to improve and sustain national systems. Having the government carry out the verification can be cheaper and can build state capacity, but once again there is an even greater conflict of interest since it would have an incentive to say that the results have been achieved even if this

186 Cordaid (2014) and Results for Development (2016)

187 Cordaid (2014)

188 Lurton (2018)

189 Results for Development (2016)

190 Results for Development (2016)

191 Results for Development (2016) and Cambridge Education (2015)

**Box 3: Perverse Incentives, Gaming, and Corruption During the Implementation of RBF Projects**

It is not controversial to say that all types of development financing, RBF or not, are susceptible to perverse incentives, gaming, and corruption. The question is whether there are more instances of these kinds of behavior in RBF projects than in traditionally financed projects.

The International Anti-Corruption Resource Center (IACRC) has estimated that private sector corruption in traditional development financing has cost developing countries over US$500 billion (IACRC, 2018). In response, the IACRC has issued guidance on how to spot the biggest red flags in development projects, including bid rigging, collusive bidding, and fraud. These are mostly related to procurement and tend to be in the form of inflated invoices or outright fraudulent transactions as was the case in the infamous Cashgate scandal in Malawi (The Economist, 2014).

Broadly speaking, procurement is often not a part of RBF projects, or rather, procurement does not follow donor agency rules (another reason why country systems are needed for RBF to work). Thus, there has not been any definitive evidence of such corruption in RBF projects, but our survey interviewees talked about how these projects can create unanticipated perverse incentives to game data. In Cameroon, one development agency staff member noticed that, when the verification team made a school visit and asked the headteacher for a document, the headteacher said that she had to look for it in her office and came back 20 minutes later with a brand new, fake paper. This is one of the reasons why teams often insist on independent, third-party verification. Another perverse incentive that RBF can introduce is to move the focus of actors from the long term to the short term (Bond, 2017). This was the case in the GEC, where small NGOs ended up prioritizing specific targets rather than working to develop sustainable, long-term gains.

is not the case. The third option is to use an independent verification agent such as a local NGO or an audit firm. While this can ensure impartiality, it can also lead to higher costs, both in time and money. In fragile and conflict states, there are additional issues. First, there may not be any reliable local organizations able to carry out the verification, and second, as a recent policy note on RBF experiences in Niger has suggested, the procurement processes to hire verification firms can be very lengthy, which would delay implementation and verification.[192] Conflicts of interest can also exist in these situations, though these will manifest differently than in the case of governments. In countries that lack sufficient capacity to collect and analyze data, it is likely that there will be a limited number of actors available for verification (for example, from a large research university) to provide both technical guidance to the government and independent verification. Since verification agents need to be independent, which usually means that they will not be eligible to help to implement other project activities. For example, in Nepal, civil service organizations were upset that the University of Kathmandu was going to be the independent verification agent for a GPE grant because that meant that the University of Kathmandu could not provide technical guidance on the program. Per the World Bank DLI analysis, 202 DLIs out of 352 (about 57 percent) required third party verification, while roughly 28 percent relied on

192 Majgaard et al (2018)

national institutions for verification, e.g. the State Examination Board or General Auditor. The remaining DLIs were to be verified by the World Bank, which generally involves World Bank teams checking nationally produced data (e.g. Ministry of Education documents or status reports).

In practice, many teams will opt for third party verification initially as a way to build capacity (usually by pairing the firm with a local agency) and then transfer responsibility for the verification process to national agencies subsequently. This means a high initial investment, as described previously, but for a worthwhile pay-off. For example, in Sierra Leone, the team used an independent, third-party firm for verification in year 1, which was very expensive and still involved some quality issues, but are now using government systems in year 2, which has required revising and strengthening protocols (such as rotating enumerators every quarter, sending multiple enumerators to the same school per quarter (sample based), and having district officers make spot checks at schools). This tiered approach has also been adopted by a project in Nepal, where the disbursement of funds was conditional on an independent audit of the EMIS, thus giving the government an incentive to have the system evaluated and to make the necessary improvements so that future verification would be more reliable.

There is no comprehensive overview of the relative costs of verification mechanisms, but there is some evidence that these costs vary a lot. A recent study has calculated verification costs as a fraction of total program costs for health interventions, with a 2011 intervention in Burundi having verification costs equal to 1 percent of the project's total cost, whereas Plan Nacer having spent around 10 percent of its maximum bonus payments on verification. A program in Benin paid as much as 50 percent of its bonus payments on verification.[193] Therefore, context and, more importantly, the *verification mechanism* matter.

How effective is traditional verification that relies on human supervision (usually by the donor, a third party, or the government)? Unfortunately, this sort of verification can often fail to identify misreporting, even if the percentage of data supervised/checked is large.[194] Thus, unless it is widespread, misreporting could go unnoticed. This is a troubling finding since many RBF interventions rely on these spot checks.

A recent verification system that has been garnering some interest is automated (or algorithmic) monitoring in which machine learning and algorithmic techniques are used to pinpoint possible problems in the data. The hope is that using this approach would reduce verification costs, for example, because the algorithm will come up with a list of schools, clinics, or other fund recipients that may be misreporting data.

Researchers tried this method out on data from a health RBF intervention in Zambia.[195] They tested several machine learning algorithms to see whether any of them could successfully predict which clinics were over-reporting data. For this, they used a sample of reported and verified data from 140 clinics. What they found was that some algorithms[196] were quite successful at identifying which clinics are likely to be misreporting data. An additional benefit is that these algorithms should get increasingly accurate as more data are entered. They can also be used to increase the accuracy and reliability of existing administrative datasets.

Another methodology using a series of algorithmic simulations has been proposed as a way to pinpoint potential data issues and has been tested using data from a

193 Grover et al (2018)

194 Grover et al (2018)

195 Grover et al (2018)

196 Specifically, the random forest algorithm, see paper for more details.

health intervention in Benin.[197] The results of this experiment showed that using this methodology could bring verification costs down from 30 percent to around 20 percent. Thus, automated methods are a promising possibility for bringing down the costs of verification in RBF projects. However, the evidence so far is mostly based on simulations using existing data.

### Gaming and Cheating

Even the most well-designed information and monitoring system can be prone to gaming and cheating. Any indicator that is used to distribute funding *de facto* becomes a high stakes indicator.

Several factors are likely to increase the likelihood of strategic or gaming behavior happening in education interventions: (i) the higher the stakes are; (ii) the longer the intervention has been operating; and (iii) the less able the agents or recipients are to influence the actual outcome.[198] These make intuitive sense. If an indicator determines whether or not agents receive a large sum of money, they are more likely to report having achieved it regardless of whether or not this is actually the case. Similarly, the more helpless agents feel in terms of controlling the results, the more likely they are to game it. If agents have no control over an outcome that is being used as a metric, this defeats the purpose of RBF.

One way to evaluate whether the results have really been achieved is to use complementary indicators. For example, an intervention in Kenya rewarded teachers for improving students' test scores in the government exam, and these subsequently were seen to have improved. So it seemed as though the program had led to improvements in student learning. However, when the students took a complementary test that measured the same skills, the improvement disappeared. It seemed that the teachers had focused on teaching kids how to take the government exam rather than on developing their skills and improving the teaching content as activities such as homework and teacher attendance had remained unchanged, but the number of exam preparation sessions had increased.[199]

There may be several ways to get around this. One way, as previously discussed, might be to distribute the funding upon the achievement of a range of indicators rather than just one or two. This would relieve the pressure on the recipient as they would receive funding even if they only achieved some but not all of the indicators. One example of an intervention that used several indicators was the Big Results Now program in Tanzania. The only final outcome indicator for education (linked to disbursements) was an improvement in students' reading, writing, and arithmetic skills, but it only represented 13 percent of the total budget.[200] Also, to minimize gaming, the final outcome was measured using a new test of reading, writing, and arithmetic implemented in only a random sample of schools. The program's designers argued that this would decrease the risk of gaming because it would not give teachers and schools a chance to prepare students for the test in advance. However, this did not prevent gaming in other areas of the program. Since increasing passing rates was one of the objectives of Big Results Now, many schools prevented struggling students from taking the exam in order to improve their metrics.[201]

There is not a lot of concrete evidence of perverse incentives when it comes to RBF activities in the health sector, though that

197 Lurton (2018)

198 Murnane and Ganimian (2014)

199 Kremer et al (2010) cited in Murnane and Ganimian (2014)

200 Janus and Kejzer (2015)

201 Cilliers et al (2018)

might be due to the fact that the monitoring of perverse effects is usually not built into the design of these projects.[202] More analysis has been done on the negative effects of conditional cash transfers (CCTs), which include cherry picking or cream skimming where governments and/or organizations target those populations that are more likely to achieve the targets, which can increase inequity. In our survey interviews, practitioners mentioned that the most common form of gaming was the tendency for clients to make no effort to achieve non-remunerated indicators.

In general, there is no strong evidence of an overwhelming number of negative unintended consequences associated with RBF, though teams may not be establishing the monitoring mechanisms needed to catch such undesirable behavior.

In our interviews with practitioners, some seemed to accept the inevitability of gaming but emphasized that teams should first observe the perverse behavior in action before re-designing parts of the project to minimize them.

### Implementation Quality

In RBF projects, there is a lot of emphasis on design. Many researchers and practitioners alike will insist that the design of the incentives will make or break the project. However, even if a project is well-designed, implementation is never easy to predict, yet it is exactly *how* indicators are achieved that really matters for long-term impact. Donors often seek to control how an indicator is achieved by putting in place verification protocols, which are steps that must be checked to validate that an indicator has been achieved.

One of the issues with verification protocols in RBF projects is that they can become mechanical, quantitative exercises, which defeats the purpose of ensuring quality implementation. Quality implementation is difficult to explicitly define, but indicators should achieved in a meaningful way, e.g. not just counting the number of teachers who are trained but ensuring that they apply what they learned in the classroom.

The easy solution might seem to be to simply design indicators better or to create stringent, quality-oriented protocols, but in the end, the indicator still must be measurable (and thus, quantifiable). Also, it is important to state that verification cannot, and does not, dictate every action that a relevant actor takes to achieve that indicator. In this regard, the examples used in this section are meant to show the realities of implementation through the lens of verification, which is what many practitioners spoke about when asked in our interviews about how they ensured quality implementation, perhaps because it is more documented than other facets of implementation.

For example, one of the very first Program for Results at the World Bank was the Bridges Improvement and Maintenance Program in Nepal. To date, this is one of two PforRs at the World Bank that are now closed and is the only one for which an implementation completion report is publicly available. In the completion report, the team wrote about the challenges that they faced while implementing the program, namely, that quality was difficult to achieve. One of the DLIs was "new bridges built or improved on," which was measured by the length in meters of the bridges that had been built or improved. The completion report indicated that poor construction quality was a constant issue and that the National Planning Commission had recommended that disbursements should not be made for four new bridges because of their poor quality. The World Bank team then had to work closely with the Department of Roads (DoR) to improve the quality of construction quality because the DoR staff did not have the requisite capacity to carry out the quality assurance.[203] To do

202 Grittner (2013)

203 World Bank (2018b)

this, the team had to add a technical assistance component to the project.

These types of quality issues are not unique to infrastructure and are often even more difficult to address in the education sector. In Ghana, a project called for the distribution of iBoxes, which were devices on which were uploaded video lessons, interactive applications, and video tests as well as over 3,000 additional open source resources. One of the project's DLIs was to ensure that teachers received training in how to use the iBox and that the iBoxes were functional. While this DLI was independently verified, the team discovered that the government had not taken into account the quality of the results or their sustainability. In other words, in many instances, the iBoxes are not being used in classrooms, even though the teachers have completed the training.

An example that illustrates the difficulty of balancing stringency and quality control in protocol design is in Uganda, a project with GPE funds. The project team very carefully thought through the protocols to be included and wanted not only to train the teachers but also to ensure that there was evidence that the teachers had applied some of what they learned in their training. They came up with a four-step verification protocol that required proof that: (i) a training schedule existed; (ii) teachers had been given initial training; (iii) teachers had participated in supervision/support meetings with a coach; and (iv) teachers had received training materials. As the verification process was underway, the team learned that it was very difficult to track whether or not teachers had met with their coaches because the coaches often forgot to sign a log indicating that they had met with the teachers.

In the end, the team saw this as a lesson learned and an opportunity to stress the importance of documentation with teachers and their coaches and were able to continue monitoring all four steps while making disbursements when only three were met. There needed to be a balance between ensuring quality and allowing for some flexibility during implementation.

Generally speaking, protocols that are developed jointly by the project team and the client can improve the way in which results are achieved as well as provide the necessary flexibility to make adjustments when needed as has been necessary in several projects (in Moldova and the Dominican Republic, for example). In addition, many teams have incorporated both top-down and bottom-up accountability measures. In some performance-based school grant schemes, schools must submit improvement plans and/or evidence that they have used the money for eligible items to the provincial level, while the community also must vouch that the school received the funds. For example, under the Indonesia performance grant program funded by REACH, schools were required to spend the grant money either on teacher training or to purchase learning equipment.

In general, high quality implementation requires technical assistance. Every one of our interviewees stressed the importance of working with their counterparts to help them to achieve the indicators. For example, the ADB has built a "review and corrective action" mechanism into their Sri Lanka Education Sector Development Program, which enabled them to make changes to the design whenever they were faced with challenges without having to halt their pilot programs. This feedback loop of identifying problems and then correcting the course of the program helped the team to work more effectively with implementing agencies to achieve better results. This is known as adaptive implementation.

Another survey respondent discussed a similar situation in Lebanon. "In Lebanon, we have invested in parallel technical assistance support to the Ministry of Education to ensure that comprehensive planning takes place that focuses both on how to achieve the DLIs but also how to strengthen the system as a whole." Experience in the health sector has shown

the importance of adaptive implementation, with adaptation and change during a project's implementation being "the norm" rather than the exception" in the sector.[204] RBF can enable flexible course correction since it involves the setting of targets but does not dictate the way in which to achieve them.

Investing in technical assistance often adds to the cost of RBF activity, but it is often necessary for successful and meaningful implementation, especially in the many countries that are still familiarizing themselves with RBF.

As mentioned above, adaptive implementation means building flexibility into the design of RBF projects to enable teams to correct course when necessary and, therefore, it must be context-specific. While there are fewer studies on implementation in the education sector than in the health sector, the literature in health has confirmed the importance of customizing project design—on what is called "artisanal RBF"[205]—and of disseminating information about RBF. For example, in an intervention in Nigeria, it was found that part of the variation in performance within a single RBF scheme was due to implementation factors such as how much the implementing agents understood about RBF or how well front-line agents communicated with different government levels and vice versa.[206]

In a recent study on RBF in the health sector in fragile and conflict states, a team of researchers found that most successful interventions had had to adapt their original plans and come up with local innovations in order to improve delivery.[207] A program in South Kivu (in the Democratic Republic of Congo) included mechanisms such as non-performance-based payments to jumpstart operations (*bonus de demarrage*) and harnessing the support of the community for the rehabilitation of health facilities, whereas in Adamawa State (Nigeria), regular monthly meetings were set up between donors, local authorities, and implementers to improve coordination and avoid conflicts. These were *ad hoc* solutions to problems that were not discovered until after implementation had started.

With RBF projects, there is usually no prescribed implementation plan in place, and it is difficult to predict what challenges are likely to be encountered. As in the Nigeria example above, sometimes working groups have been set up to come up with good practices and solutions to implementation problems.[208] This can also be done at the central level. Some of these solutions may be idiosyncratic and unique to a particular national context. For example, Rwanda introduced community verification of DLIs as a tool, while Burundi tested out the complementarities of reducing user fees and implementing RBF at the same time in healthcare delivery.[209] The key is to understand that RBF is not a static modality or blueprint but a process of continued improvement.[210]

204 Ridde et al (2018)
205 World Bank (2017)
206 Ma-Nitu et al (2018)
207 Bertone et al (2018)
208 Ma-Nitu et al (2018)
209 Ma-Nitu et al (2018)
210 Ma-Nitu et al (2018), Bertone et al (2018)

## Failure to Achieve Targets

There is not much evidence on how often recipients fail to achieve their targets and what the repercussions are. One of the critiques of RBF is that disbursements will happen regardless of results because it is too politically difficult to withhold funds. In many cases, development agency teams simply change the indicators (60 percent of survey respondents indicated that was what they had done). In fairness, even non-RBF projects end up changing indicators, and teams argue that they revise indicators to reflect changing priorities and realities on the ground and not based on their whims.

There is some concrete evidence that teams have taken measures to either reduce or withhold funds when they felt that things were not going well (54.3 percent of respondents indicated that they have done just that). In the Burundi Common Education Fund (2011-2015), the Bank project team reduced transfers to some schools and meso-level institutions if their internal controls and accounting had been under-par in the previous year.

Out of the 51 results-based projects in the World Bank portfolio, only six have closed, and only four have completion reports. Of the four, two projects in Jamaica did not disburse the full loan amount because the government had miscalculated the total amount of loan that it required. The completion reports for the other two projects, in Bangladesh and Pakistan, indicated that the total commitment amount was not disbursed due to unmet DLIs. In the Bangladesh Primary Education Project, the team revised the indicators several times and ended up canceling several unmet DLIs (US$8.3 million out of the total US$100 million was canceled).[211] In the Pakistan Tertiary Education Support project, US$77.82 million of the total US$222.18 million was cancelled as a result of the non-achievement of disbursement-linked indicators. In the completion report, the independent auditor indicated that, although the team had lowered the DLI targets using a "thin" justification, the new targets helped to move the project forward, and the attainment of the outputs suggested that there was substantial improvement in learning conditions.[212]

During our interviews with practitioners, there were some who reported that some canceled funds had simply been reallocated to a non-RBF portion of the project. This was not the case in Bangladesh or Pakistan, and there is no substantial evidence to corroborate that observation, but further analysis could be done to investigate how often funds are genuinely canceled as opposed to redistributed.

For projects that have not yet closed, disbursement flexibility is still a key feature, with one respondent writing, "we are withholding funds for now as many of our DLIs have a roll-over option."

Undoubtedly, all stakeholders want to have successful projects, regardless of RBF, and are likely to do what is in their power to ensure a positive outcome. There is a fine line between incentivizing and rewarding the effort that goes into achieving results and signaling to country governments that there are no real consequences if indicators are not met. If the latter occurs, then the promise of RBF is minimized. As emphasized previously, if information about RBF is widely disseminated upfront, then the bad news of withholding funds based on non-performance should not be unexpected and thus be somewhat more palatable.

211 World Bank (2018c)

212 World Bank (2018d)

## Sustainability

An area where questions remain, and where there is little research available, is the sustainability of RBF activities. Generally, the sustainability of results seems promising for areas such as conditional cash transfers. For example, children who experienced PROGRESA/Oportunidades in Mexico have been shown to have better education and labor market outcomes years after the program (though for other transfers the effects have faded out).[213/214] However, for interventions such as teacher incentives or performance grants, the evidence seem to be lacking, to a great extent because most are not designed to measure their long-term effects.

In other areas such as health, there have been several examples that suggest that incentives can have a lasting effect, particularly a randomized control trial in Argentina conducted with Plan Nacer and the previously mentioned Cordaid intervention in Uganda.

Plan Nacer is a well-known program that provides health insurance to otherwise uninsured pregnant women and children. In a field experiment, the program randomly paid a 200 percent premium to treatment clinics that initiated prenatal care, and a study of this intervention found that those clinics that were paid the premium had a 34 percent increased rate of initiation of services than those who were not paid the premium.[215] The study found that these higher levels of initiation of prenatal care persisted for at least 18 months, and in some cases 24 months, after the incentives ended. This evidence speaks to the fact that temporary incentives can produce longer-term behavioral changes without creating unsustainable financing expectations. Plan Nacer is unusual in that it is a long-running government-sponsored program that was originally funded by the World Bank and the central government with the provincial governments taking on responsibility for 30 percent of the total costs later on.[216] Also, the central government was already financially solvent, which is not always the case for all countries.[217]

213 Behrman et al (2009)

214 Parker and Vogl (2018)

215 Celay et al (2015)

216 Center for Global Development (2018)

217 Berman (2015)

# Concluding Remarks

**Currently, there is more robust evidence available on how RBF can have an impact on education at the level of specific groups and individuals (such as teachers) and less evidence available on its effectiveness at the national or programmatic level.** However, in some ways, aligning donors and clients with the aim of prioritizing the achievement of results, whether they are inputs or processes or outcomes, is a legitimate goal in and of itself. Unsurprisingly, RBF cannot serve as a substitute for a strong theory of change, nor can it compensate for improperly identifying the types of binding constraints in an education system that can be unlocked by incentives.

In particular, international development will forever be anchored in country context, and until there are more evaluations of RBF and governments in education, it will be difficult to make broader statements about its effectiveness. The evidence base continues to grow as more new research comes out (for example, REACH is funding three country-level assessments of RBF as well as a round of proposals that focus on incentives at the district/regional/provincial level).

To improve the odds of lasting change, design and implementation matter a lot, and RBF often requires even more planning than a traditionally financed project. For RBF to work well, stakeholders must genuinely think through the results chain, and not fall into the habit of identifying disparate activities that could be financed. RBF forces donors and countries alike to think about how those different activities interact in the education system, and which behaviors might respond to incentives.

To this end, evidence thus far shows:

1. **RBF and teachers:** Teacher incentives can but do not always improve teacher attendance and student learning. The design of the incentive scheme and the context matter. The effects are larger and more positive in developing country contexts.
2. **RBF and students and families:** Student and family incentives (such as CCTs, for instance) has a good track record of reducing school dropout and increasing school attendance, though the evidence for its effects on student learning is more mixed. Conditional transfers to students tied to their own learning are a promising area of future research.
3. **RBF and schools:** The evidence base on the effectiveness of performance-based grants is still quite limited. For now, it seems that in some cases they can work, especially when grants are combined with other interventions such as capacity building (for example, to principals and school committees) or when money is spent on inputs that affect learning outcomes.
4. **Synergies:** There is growing evidence that combining different RBF interventions within the same program can generate results that go beyond the sum of any two interventions alone. Though the research is limited, this suggests that RBF that tackles several bottlenecks at once can have larger effects.

For RBF and Governments, several key criteria are critical for more effective RBF:

1. Choosing RBF as the appropriate financing modality requires careful consideration of political commitment, and understanding the risks involved, costs, and country context (for example, capacity and country systems).
2. RBF project design should prioritize the cascading of incentives and should select and price indicators with an objective or methodology in mind. Some of these include cost effectiveness, increasing the chances of achieving other indicators, or reducing risk of nonpayment.
3. RBF project implementation should think of the purpose of monitoring and information systems, invest upfront in verification, and be adaptive and flexible in order to address realities on-the-ground.

Ultimately, there is proof that RBF can have a positive impact on learning conditions and, in rare instances, on increasing learning itself. This makes it a powerful financing modality for policymakers around the world to consider using in their education sectors.

# References


ADB (2013). Education Sector Development Program (2013-2017). Disbursement Linked Results and Indicators.

ADB (2014). Supporting Kerala's Additional Skill Acquisition Program in Post-Basic Education. Program Fiduciary Systems Assessment.

ADB (2016). Midterm Review of Results-Based Lending for Programs.

ADB (2017). Results-Based Lending at the Asian Development Bank: An Early Assessment.

Akresh, R., De Walque, D., & Kazianga, H. (2013). *Cash transfers and child schooling: evidence from a randomized evaluation of the role of conditionality.* The World Bank.

Al-Samarrai, S., Shrestha, U., Hasan, A., Nakajima, N., Santoso, S., & Wijoyo, W. H. A. (2018). Introducing a performance-based component into Jakarta's school grants: What do we know about its impact after three years?. *Economics of Education Review*, *67*, 110-136.

Andrews, M. (2013). *The limits of institutional reform in development: Changing rules for realistic solutions*. Cambridge University Press.

Andrews, M., Pritchett, L., & Woolcock, M. (2013). Escaping capability traps through problem driven iterative adaptation (PDIA). *World Development*, *51*, 234-244.

Andrews, M., Woolcock, M., & Pritchett, L. (2017). *Building state capability: Evidence, analysis, action*. Oxford University Press.

Ashraf, N. (2009). Spousal control and intra-household decision making: An experimental study in the Philippines. *American Economic Review*, *99*(4), 1245-77.

Ashraf, N., Bandiera, O., & Lee, S. S. (2014). Do-gooders and go-getters: career incentives, selection, and performance in public service delivery. *STICERD-Economic Organisation and Public Policy Discussion Papers Series*, *54*.

Baird, S., McIntosh, C., & Özler, B. (2011). Cash or condition? Evidence from a cash transfer experiment. *The Quarterly Journal of Economics*, *126*(4), 1709-1753.

Baird, S., Ferreira, F. H., Özler, B., & Woolcock, M. (2014). Conditional, unconditional and everything in between: a systematic review of the effects of cash transfer programmes on schooling outcomes. *Journal of Development Effectiveness*, *6*(1), 1-43.

Barlevy, G. & Neal, D. (2012). Pay for percentile. *American Economic Review*, *102*(5), 1805-31.

Barón, J. & Adelman, M. (2018). HAITI: Can Pre-Conditions for RBF be Established in Fragile States? REACH Policy Note.

Barrera-Osorio, F. & Raju, D. (2015). Teacher performance pay: Experimental evidence from Pakistan.

Barrera-Osorio, F. & Ganimian, A.J. (2016). The barking dog that bites: Test score volatility and school rankings in Punjab, Pakistan. *International Journal of Educational Development*, *49*, 31-54.

Barrera-Osorio, F., Bertrand, M., Linden, L. L., & Perez-Calle, F. (2011). Improving the design of conditional transfer programs: Evidence from a randomized education experiment in Colombia. *American Economic Journal: Applied Economics*, *3*(2), 167-95.

BBC (2018). Guinea country profile. https://www.bbc.com/news/world-africa-13442051. Accessed July 2018.

Barham, T., Macours, K., & Maluccio, J.A. (2013). *More schooling and more learning? Effects of a three-year conditional cash transfer program in Nicaragua after 10 years* (No. IDB-WP-432). IDB Working Paper Series.

Behrman, J.R., Parker, S.W., & Todd, P.E. (2009). Schooling impacts of conditional cash transfers on young children: Evidence from Mexico. *Economic development and cultural change*, *57*(3), 439-477.

Behrman, J.R., Parker, S.W., Todd, P.E., & Wolpin, K.I. (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.

Berman, D. (2015). Argentina: can short term incentives change long term behavior?

Bernal, P., Celhay, P., & Martinez, S. (2018). *Is results-based aid more effective than conventional aid? Evidence from the health sector in El Salvador* (No. IDB-WP-859). IDB Working Paper Series.

Berry, J. (2015). Child Control in Education Decisions: An Evaluation of Targeted Incentives to Learn in India. *Journal of Human Resources*, *50*(4), 1051-1080.

Bertone, M. P., Jacobs, E., Toonen, J., Akwataghibe, N., & Witter, S. (2018). Performance-based financing in three humanitarian settings: principles and pragmatism. *Conflict and health*, *12*(1), 28.

Birdsall, N. & Barder, O. (2006). *Payments for Progress: A Hands-Off Approach to Foreign Aid* (No. 102).

Birdsall, N. & Savedoff, W.D. (2012). Cash on delivery: a new approach to foreign aid. CGD Books.

Blimpo, M. P. (2014). Team incentives for education in developing countries: A randomized field experiment in Benin. *American Economic Journal: Applied Economics*, *6*(4), 90-109.

Blimpo, M.P., Evans, D.K., & Lahire, N. (2015). *Parental human capital and effective school management: evidence from The Gambia*. The World Bank.

Bloom, N., Eifert, B., Mahajan, A., McKenzie, D., & Roberts, J. (2013). Does management matter? Evidence from India. *The Quarterly Journal of Economics*, *128*(1), 1-51.

Bloom, N., Lemos, R., Sadun, R., & Van Reenen, J. (2015). Does management matter in schools? *The Economic Journal*, *125*(584), 647-674.

Bond (2017). Does 'skin in the game' improve the level of play? https://www.bond.org.uk/resources/does-skin-in-the-game-improve-the-level-of-play

Bowles, S. & Polania-Reyes, S. (2012). Economic incentives and social preferences: substitutes or complements? *Journal of Economic Literature*, *50*(2), 368-425.

Cambridge Education (2015). Evaluation of the Pilot Project of Results-Based Aid in the Education Sector in Ethiopia.

Carneiro, P.M., Koussihouèdé, O., Lahire, N., & Meghir, C. (2016). School grants and education quality: experimental evidence from Senegal.

Celhay, P., Gertler, P., Giovagnoli, P., & Vermeersch, C. (2015). *Long-run effects of temporary incentives on medical care productivity.* The World Bank.

Center for Global Development (2018). Argentina's Plan Nacer. Paying for Provincial Performance in Health. http://millionssaved.cgdev.org/case-studies/argentinas-plan-nacer

Cerdán-Infantes, P. (2018). Developing an Effective Management and Information System for Education Quality in Colombia.

Cilliers, J., Kasirye, I., Leaver, C., Serneels, P., & Zeitlin, A. (2013). Improving teacher attendance using a locally managed monitoring scheme: Evidence from Ugandan Primary Schools. *Rapid response paper for International Growth Centre*.

Cilliers, J., Mbiti, I., & Zeitlin, A. (2018). Accountability and school performance: Evidence from Big Results Now in Tanzania. RISE Conference.

Clist, P. (2016). Payment by results In development aid: All that glitters is not gold. *The World Bank Research Observer*, *31*(2), 290-319.

Clist, P. (2018). Payment by results in international development: Evidence from the first decade. *Development Policy Review*.

Clist, P. & Dercon, S. (2014). 12 Principles for Payment By Results (PbR) In International Development. *DFID Research for Development. London: DFID*.

Clist, P. & Verschoor, A. (2014). The conceptual basis of payment by results. *London: DFID*.

Coffey (2016). Girls Education Challenge Process Review Report.

Cordaid (2014). Open Development Movement. Co-Creation Leads to Transformation. *Cordaid Position Paper.*

Cordaid (2017). Strengthening Health Systems through RBF.

Cruz-Aguayo, Y. & Martinez, S. (2016). *Setting Targets for Results Based Financing Programs: A Simple Cost Benefit Framework*. Inter-American Development Bank.

Das, J., Dercon, S., Habyarimana, J., Krishnan, P., Muralidharan, K., & Sundararaman, V. (2013). School Inputs, Household Substitution, and Test Scores. *American Economic Journal: Applied Economics*, 29-57.

Dauphin, A., Lahga, E., Fortin, B., & Lacroix, G. (2011). Are Children Decision-Makers within the Household? *The Economic Journal*, *121*(553), 871-903.

De Ree, J., Muralidharan, K., Pradhan, M., & Rogers, H. (2015). *Double for nothing? Experimental evidence on the impact of an unconditional teacher salary increase on student performance in Indonesia* (No. w21806). National Bureau of Economic Research.

DFID (2014). DFID's Approach to Payment by Results.

Duflo, E., Hanna, R., & Ryan, S.P. (2012). Incentives work: Getting teachers to come to school. *American Economic Review*, *102*(4), 1241-78.

Evans, D. K. & Popova, A. (2016). What really works to improve learning in developing countries? An analysis of divergent findings in systematic reviews. *The World Bank Research Observer*, *31*(2), 242-270.

Fiszbein, A. & Schady, N.R. (2009). *Conditional cash transfers: reducing present and future poverty*. The World Bank.

Fryer, R. G. (2013). Teacher incentives and student achievement: Evidence from New York City public schools. *Journal of Labor Economics*, *31*(2), 373-407.

Fryer Jr, R.G., Levitt, S.D., List, J., & Sadoff, S. (2012). *Enhancing the efficacy of teacher incentives through loss aversion: A field experiment* (No. w18237). National Bureau of Economic Research.

Galiani, S. & McEwan, P.J. (2013). The heterogeneous impact of conditional cash transfers. *Journal of Public Economics*, *103*, 85-96.

Garbers, Y. & Konradt, U. (2014). The effect of financial incentives on performance: A quantitative review of individual and team-based financial incentives. *Journal of occupational and organizational psychology*, *87*(1), 102-137.

Gelb, A., Diofasi, A., & Postel, H. (2016). Program for results: the first 35 operations. *Center for Global Development Working Paper 430.*

Gertler, P. (2004). Do conditional cash transfers improve child health? Evidence from PROGRESA's control randomized experiment. *American economic review*, *94*(2), 336-341.

Gertler, P.J., Patrinos, H.A., & Rubio-Codina, M. (2012). Empowering parents to improve education: Evidence from rural Mexico. *Journal of Development Economics*, *1*(99), 68-79.

Gilligan, D.O., Karachiwalla, N., Kasirye, I., Lucas, A., & Neal, D.A. (2018). Educator Incentives and Educational Triage in Rural Primary Schools.

Glewwe, P., Ilias, N., & Kremer, M. (2010). Teacher incentives. *American Economic Journal: Applied Economics*, *2*(3), 205-27.

Glewwe, P. & Kassouf, A.L. (2012). The impact of the Bolsa Escola/Familia conditional cash transfer program on enrollment, dropout rates and grade promotion in Brazil. *Journal of development Economics*, *97*(2), 505-517.

Glewwe, P. & Muralidharan, K. (2016). Improving education outcomes in developing countries: Evidence, knowledge gaps, and policy implications. In *Handbook of the Economics of Education* (Vol. 5, pp. 653-743). Elsevier.

Gneezy, U., Meier, S., & Rey-Biel, P. (2011). When and why incentives (don't) work to modify behavior. *Journal of Economic Perspectives*, *25*(4), 191-210.

Grittner, A.M. (2013). Results-based financing: evidence from performance-based financing in the health sector.

Grover, D., Bauhoff, S., & Friedman, J. (2018). *Using Supervised Learning to Select Audit Targets in Performance-Based Financing in Health: An Example from Zambia* (No. 481).

Holzapfel, S. & Janus, H. (2015). Improving Education Outcomes by Linking Payments to Results-An Assessment of Disbursement-Linked Indicators in Five Results-Based Approaches.

IACRC (2018). International Anti-Corruption Resource Center. https://iacrc.org/.

ICAI (2018). DFID's approach to value for money in programme and portfolio management: A performance review.

IEG (2013). ICR Review of the Pakistan Sindh Education Sector Project.

IEG (2016). Program-for-Results. An Early-Stage Assessment of the Process and Effects of a New Lending Instrument.

Imberman, S.A. (2015). How effective are financial incentives for teachers? *IZA World of Labor.*

Islam, A., Kwon, S., Masoon, E., Prakash, N., & Sabarwal, S. (2017). Information before Incentives: Goals, Expectations, and Effort among Zanzibari secondary students.

Janus, H. & Keijzer, N. (2015). Big results now? Emerging lessons from results-based aid in Tanzania.

Kerwin, J. & Thornton, R.L. (2018). Making the Grade: The Sensitivity of Education Program Effectiveness to Input Choices and Outcome Measures.

Kim, J.S. & Sunderman, G.L. (2005). Measuring academic proficiency under the No Child Left Behind Act: Implications for educational equity. *Educational Researcher*, *34*(8), 3-13.

Klingebiel, S., Gonsior, V., Jakobs, F., & Nikitka, M. (2016). *Public Sector Performance and Development Cooperation in Rwanda.* https://www.researchgate.net/publication/311256916_Imihigo_and_Development_Cooperation_What_Kind_of_Relationship.

Kremer, M., Glewwe, P., & Ilias, N. (2010). Teacher incentives. *American Economic Journal: Applied Economics*, *2*(3).

Lavy, V. (2002). Evaluating the effect of teachers' group performance incentives on pupil achievement. *Journal of political Economy*, *110*(6), 1286-1317.

Lavy, V. (2009). Performance pay and teachers' effort, productivity, and grading ethics. *American Economic Review*, *99*(5), 1979-2011.

Lazear, E.P. (2003). Teacher incentives. *Swedish Economic Policy Review*, *10*(2), 179-214.

Loyalka, P.K., Sylvia, S., Liu, C., Chu, J., & Shi, Y. (2016). Pay by Design: Teacher Performance Pay Design and the Distribution of Student Achievement.

Lurton, G. (2018). First approach to algorithmic RBF supervision. https://grlurton.github.io/orbf_data_validation/Analysis.html

Ma-Nitu, S. M., Tembey, L., Bigirimana, E., Dossouvi, C. Y., Basenya, O., Mago, E., Salongo, P.M., Zongo, A., & Verinumbe, F. (2018). Towards constructive rethinking of PBF: perspectives of implementers in sub-Saharan Africa. *BMJ Global Health*, *3*(5), e001036.

Majgaard, K., Ouedraogo, A., Mallberg, M., Andriamarofara, H., & Mulet, P. (2018). Early lessons around results-based aid for education in Niger. *REACH Policy Note.*

Marsh, D.R., Schroeder, D.G., Dearden, K.A., Sternin, J., & Sternin, M. (2004). The power of positive deviance. *Bmj*, *329*(7475), 1177-1179.

Mbiti, I.M. (2016). The need for accountability in education in developing countries. *Journal of Economic Perspectives*, *30*(3), 109-32.

Mbiti, I., Muralidharan, K., Romero, M., Schipper, Y., Manda, C., & Rajani, R. (2018). *Inputs, Incentives, and Complementarities in Education: Experimental Evidence from Tanzania* (No. w24876). National Bureau of Economic Research.

Mbiti, I., Romero, M., & Schipper, Y. (2018b). *Left Behind by Optimal Design: The Challenge of Designing Effective Teacher Performance Pay Programs*. Technical report, working paper.

McEwan, P.J. (2015). Improving learning in primary schools of developing countries: A meta-analysis of randomized experiments. *Review of Educational Research*, *85*(3), 353-394.

Molina-Millan, T., Barham, T., Macours, K., Maluccio, J.A., & Stampini, M. (2016). *Long-term impacts of conditional cash transfers in Latin America: Review of the evidence*. Inter-American Development Bank.

Muralidharan, K. (2012). Long-Term Effects of Teacher Performance Pay: Experimental Evidence from India. *Society for Research on Educational Effectiveness*.

Muralidharan, K. & Sundararaman, V. (2011). Teacher performance pay: Experimental evidence from India. *Journal of Political Economy*, *119*(1), 39-77.

Murnane, R.J. & Ganimian, A.J. (2014). Improving educational outcomes in developing countries: Lessons from rigorous evaluations. Cambridge, MA: National Bureau of Economic Research.

Olander, S. & Högberg, A.S. (2016). Practical Guidance for Results-Based Financing Approaches. SIDA.

Parker, S.W. & Vogl, T.S. (2018). Long-term effects of cash transfers: Looking to the next generation. https://voxeu.org/article/long-term-effects-cash-transfers.

Paul, E., Albert, L., Bisala, B. N. S., Bodson, O., Bonnet, E., Bossyns, P., ... & Gyselinck, K. (2018). Performance-based financing in low-income and middle-income countries: isn't it time for a rethink?. *BMJ Global Health*, *3*(1), e000664.

Pearson, M., Johnson, M., & Ellison, R. (2010). Review of major Results Based Aid (RBA) and Results Based Financing (RBF) schemes. Final report. *DfID Human Development Resource Centre*.

Pritchett, L. (2001). Where has all the education gone? *The World Bank Economic Review*, *15*(3), 367-391.

Results for Development (2016). Paying for Performance: An Analysis of Output-Based Aid in Education.

RBF Health (2017). Pricing and The Use of Data in RBF: Towards a Higher Return on Investment. https://www.rbfhealth.org/blog/pricing-and-use-data-rbf-towards-higher-return-investment

Ridde, V., Yaogo, M., Zongo, S., Somé, P.A., & Turcotte-Tremblay, A.M. (2018). *Twelve months of implementation of health care performance-based financing in Burkina Faso: A qualitative multiple case study.* The International Journal of Health Planning and Management, 33(1), e153-e167.

Rusa, L. & Fritsche, G. (2007). Rwanda: performance-based financing in health. *Emerging Good Practice in Managing for Development Results: Sourcebook*-, 55-60.

Rusa, L., Schneidman, M., Fritsche, G., & Musango, L. (2009). Rwanda: Performance-based financing in the public sector. *Performance incentives for global health: potential and pitfalls*.

Sabarwal, S., Joshi, A., & Blackmon, W. (2016). A Review of the Multi-Donor Results-Based Financing mechanism used for Tanzania's Big Results Now in Education Program.

Shepard, D., Zeng, W., & Nguyen, H.T.H. (2015). Cost-Effectiveness Analysis of Results-Based Financing Programs.

Snilstveit, B., Stevenson, J., Phillips, D., Vojtkova, M., Gallagher, E., Schmidt, T., Jobse, H., Geelen, M., Pastorello, M.G., & Eyers, J. (2015). Interventions for improving learning outcomes and access to education in low-and middle-income countries: a systematic review. *3ie*.

The Economist (2014). The $32m heist. https://www.economist.com/baobab/2014/02/27/the-32m-heist

UNESCO (2018). *Walk before you run: the challenges of results-based payments in aid to education*. GEMR Policy Paper 33.

Upper Quartile (2014). Evaluation of Results Based Aid in Rwandan Education - 2013 Evaluation Report.

Upper Quartile (2015). Evaluation of Results Based Aid in Rwandan Education - Year Two.

World Bank (2001). Project Appraisal Document for the Guinea - Education for All Program (Phase I) Project.

World Bank (2008). Project Appraisal Document for the Jamaica Early Childhood Development Project.

World Bank (2009). Project Appraisal Document for the Pakistan Sindh Education Sector Project (SEP).

World Bank (2011). Implementation Completion and Results Report on the Bangladesh Primary Education Development Project II.

World Bank (2011b). Project Appraisal Document for the Bangladesh Third Primary Education Development Project.

World Bank (2013). Project Appraisal Document for the Pakistan Second Sindh Education Sector Project (SEP II).

World Bank (2014). Project Appraisal Document for the Tanzania Big Results Now in Education Program.

World Bank (2015). Achieving Results for Women's and Children's Health: 2015 Progress Report.

World Bank (2016). Project Appraisal Document for a Lebanon Support to Reaching All Children with Education Program for Results.

World Bank (2017). Results-Based Financing in Education: Financing Results to Strengthen Systems.

World Bank (2017b). Implementation Completion and Results Report on the India Technical Engineering Educational Quality Improvement Project II.

World Bank. (2017c). World Development Report 2018: Learning to Realize Education's Promise.

World Bank (2018). World Bank Open Data. https://data.worldbank.org/

World Bank (2018b). Implementation Completion and Results Report on the Nepal Bridges Improvement and Maintenance Program.

World Bank (2018c). Implementation Completion and Results Report on the Bangladesh Primary Education Development Program III.

World Bank (2018d). Implementation Completion Report Review for the Pakistan Tertiary Education Support Project.

World Bank (2018e). How to attract and motivate passionate public service providers. http://blogs.worldbank.org/impactevaluations/how-attract-and-motivate-passionate-public-service-providers

RESULTS-BASED FINANCING IN EDUCATION

# Learning from What Works

**The Results in Education for All Children (REACH) Trust Fund** *supports and disseminates research on the impact of results-based financing on education outcomes. The goal is to collect and build empirical evidence and operational lessons learned to help governments and development organizations design and implement the most appropriate results-based financing mechanisms for improved learning outcomes. For more information about who we are and what we do, go to* **worldbank.org/reach.**


**Education Global Practice**

**Results in Education for All Children (REACH)**

**World Bank**

1818 H Street, NW
Washington DC, 20433
USA

worldbank.org/reach
reach@worldbank.org



WORLD BANK GROUP Education

USAID FROM THE AMERICAN PEOPLE

Norad

Bundesministerium für wirtschaftliche Zusammenarbeit und Entwicklung