

# Evaluation of the Pilot Project of Results-Based Aid in the Education Sector in Ethiopia

Final Report  
EC 2004 - 2006  
September 2015

# Evaluation of the Pilot Project of Results-Based Aid in the Education Sector in Ethiopia

Final Report  
EC 2004 - 2006  
September 2015

## Issue and revision record

Revision	Date	Originator	Checker	Approver	Description
0	29/05/2015	Rich Tobin, Terry Allsop, Robin Ellison, Larry Orr, Michael Watts, Mark Pearson, Elisabetta Naborri, Abiy Nega and Endale Asefa	Jawaad Vohra	John Martin	Year 3 Final Report
1	21/08/2015	Rich Tobin, Terry Allsop, Robin Ellison, Larry Orr, Michael Watts, Mark Pearson, Elisabetta Naborri, Abiy Nega and Endale Asefa	Jawaad Vohra	John Martin	Revised Year 3 Final Report based on DFID's feedback
2	23/09/2015	Rich Tobin, Terry Allsop, Robin Ellison, Larry Orr, Michael Watts, Mark Pearson, Elisabetta Naborri, Abiy Nega and Endale Asefa	Jawaad Vohra	John Martin	Revised Year 3 Final Report based on DFID's feedback (e-mail received on 21 <sup>st</sup> September 2015)

### Information class: Standard

This document is issued for the party which commissioned it and for specific purposes connected with the above-captioned project only. It should not be relied upon by any other party or used for any other purpose.

We accept no responsibility for the consequences of this document being relied upon by any other party, or being used for any other purpose, or containing any error or omission which is due to an error or omission in data supplied to us by other parties.

This document contains confidential information and proprietary intellectual property. It should not be shown to other parties without consent from us and from the party which commissioned it.

# Contents

<b>Chapter</b>	<b>Title</b>	<b>Page</b>
	Acronyms and Abbreviations	i
	Executive Summary	ii
<b>1</b>	<b>Results-Based Aid in Ethiopia's Education Sector</b>	<b>1</b>
1.1	Results-based Aid: Theory	1
1.2	The RBA Pilot: Design	1
1.3	The RBA Pilot: Action in Response	4
1.4	The RBA Pilot: What Happened?	7
1.5	The RBA Pilot: The Evaluation	9
1.6	The Evaluation's Purpose, Methods, and Limitations	10
1.7	An Independent Evaluation	12
<b>2</b>	<b>The RBA Pilot's Results and Value Added</b>	<b>13</b>
2.1	Impacts on the Numbers of Sitters Attributable to the RBA Pilot	13
2.2	Impacts on the Numbers of Passers Attributable to the RBA Pilot	16
<b>3</b>	<b>The RBA Pilot's Contribution to Gender and Regional Equity</b>	<b>19</b>
3.1	Gender Equity	19
3.2	Regional Equity	21
3.3	Gender and Regional Equity Combined	22
3.4	Other Disparities	23
3.5	Conclusions	23
<b>4</b>	<b>The Systems Effects of the RBA Pilot</b>	<b>25</b>
4.1	Resource Allocation Outcomes	25
4.2	EGSECE Administration and the MoE's EMIS	25
4.3	Financial Management System and Resource Allocation Processes	26
4.4	Accountability	27
4.5	Corruption	27
<b>5</b>	<b>Understanding Why the Approach Did or Did Not Work</b>	<b>29</b>
5.1	What Happened	29
5.2	What Worked and Why	31
5.3	What Did Not Work and Why Not	32
5.4	Targeted Funding	33
5.5	Conclusions	34
<b>6</b>	<b>Impact on Aid Relationships</b>	<b>36</b>
<b>7</b>	<b>Value for Money</b>	<b>38</b>
7.1	Methodology	38
7.2	Cost Effectiveness of the RBA Instrument	40
7.3	Relative Cost Effectiveness of RBA with PBS and GEQIP	40

7.4	Optimal Level of Incentives _____	41
7.5	Effect of Using Alternative Baselines _____	41
7.6	Cost per Additional Net Sitter _____	44
7.7	Comparisons with Assumptions in DFID's <i>Business Case</i> _____	44
<b>8</b>	<b>Potential Unintended Consequences of the RBA Pilot</b>	<b>46</b>
<b>9</b>	<b>Summary, Key Lessons, and Recommendations</b>	<b>49</b>
9.1	Summary _____	49
9.2	Another Perspective on the RBA's Results _____	51
9.3	Two Key Lessons _____	52
9.4	Recommendations _____	53
<b>10</b>	<b>Bibliography</b>	<b>56</b>

## Tables

Table E.1:	Per Student Incentives for Sitting and Passing the EGSECE _____	ii
Table 1.1:	Incentives for Sitting and Passing the EGSECE _____	3
Table 1.2:	Expected Results of DFID's RBA Pilot in Ethiopia _____	4
Table 1.3:	Timeline for the RBA Pilot in Ethiopia _____	6
Table 1.4:	Comparison between Ethiopian, Gregorian, and RBA Pilot Calendars _____	6
Table 1.5:	EGSECE Sitters and Passers, EC 2003 to EC 2006 _____	7
Table 1.6:	RBA Rewards (£) Allocated to Regions _____	8
Table 1.7:	Evaluation Questions Addressed in This Report _____	10
Table 2.1:	Impacts, Years 1-3, Numbers of Boys and Girls Sitting for the EGSECE _____	15
Table 2.2:	Impacts, Years 1-3, Numbers of Boys and Girls Passing the EGSECE _____	17
Table 3.1:	Proposed Alternative Premiums in the Allocation Formula _____	24
Table 5.1:	Regional Approaches to RBA _____	30
Table 7.1:	Cost per Additional Net Sitter/Passer _____	44
Table 7.2:	Comparison of RBA Pilot Results with <i>Business Case</i> Assumptions _____	45
Table 8.1:	Potential Unintended Consequences of the RBA Pilot _____	46
Table 8.2:	EGSECE Pass Rates, EC 1999-2006 _____	46

## Figures

Figure 1.1:	Ethiopia's Regions _____	2
Figure 1.2:	DFID's Theory of Change for the RBA Pilot in Ethiopia _____	3
Figure 2.1:	Numbers of Grade 10 Students Sitting and Passing the EGSECE, EC 1999-2006 _____	14
Figure 3.1:	Gender Parity Indices for EGSECE Sitters and Passers _____	19
Figure 3.2:	Regional Equity: The Percentage of Sitters and Passers in Emerging Regions – Actual, Projected and Targets _____	21
Figure 3.3:	Gender Parity Indices for EGSECE Sitters and Passers in Emerging Regions (ERs) and Other Regions (Non-ERs) _____	23
Figure 7.1:	Value for Money Approach _____	39
Figure 7.2:	Estimated Reward Payments with Alternative Baselines _____	43
Figure 8.1:	GPA Scores in EC 2003, 2004, 2005, and 2006 _____	47

## Appendices

Acronyms and Abbreviations	_____
Appendix 1 Theories of Change and the RBA Pilot	_____
Appendix 2 Projected versus Actual Distribution of Reward Payments to Regions	_____
Appendix 3 ToR, Evaluation of Pilot Project of Results-Based Aid in the Education Sector – Ethiopia	_____
Appendix 4 Impact Estimation Data, Models, and Procedures	_____
Appendix 5 Background and Detailed Results on Equity	_____
Appendix 6 Regional and School-Level Activity EC 2003 – EC 2007	_____
Appendix 7 Value for Money and Methods for Calculating Cost Benefit/Cost Effectiveness	_____
Appendix 8 Assessment of the RBA Pilot Using the ICAI's Standards	_____
Bibliography	_____

## Acronyms and Abbreviations

CGD	Center for Global Development
CITS	Comparative Interrupted Time Series
DALY	Disability-Adjusted Life Year
DFID	Department for International Development
DHS	Demographic and Health Survey
DRS	Developing Regional States
EC	Ethiopian Calendar
EGSECE	Ethiopian General Secondary Education Certificate Examination
EMIS	Education Management Information System
ER	Emerging Region(s)
GC	Gregorian Calendar
GEQIP	General Education Quality Improvement Programme
GER	Gross Enrolment Ratio
GoE	Government of Ethiopia
GPA	Grade point average
GPI	Gender Parity Index
ITS	Interrupted Time Series
MDE	Minimum Detectable Effect
MoE	Ministry of Education
MoFED	Ministry of Finance and Economic Development
MoU	Memorandum of Understanding
NEAEA	National Educational Assessment and Examinations Agency
Non-ER	Non-Emerging Region(s)
NPV	Net Present Value
PBS	Protection of Basic Services
PTSA	Parent Teacher Student Association
RBA	Results-Based Aid
REB	Regional Education Bureau
RPI	Regional Parity Index
SNNPR	Southern Nations, Nationalities, and Peoples' Region
ToC	Theory of Change
ToR	Terms of Reference
VfM	Value for Money
WEO	Woreda Education Office
ZEO	Zonal Education Office

# Executive Summary

This report presents the results of an independent, three-year evaluation of a pilot project intended to improve access to and the quality of lower secondary education in Ethiopia through the use of results-based aid (RBA), an innovative approach to development. RBA is an aid partnership between a donor and a recipient government in which the disbursement of aid is tied to results achieved rather than activities completed or outputs produced. The amount of aid provided is directly related to the magnitude of the outcomes achieved.

Donors take a ‘hands-off’ approach and do not direct or specify how a project should be implemented or desired results achieved. Those decisions are left to the recipient. Consistent with the Paris Declaration on Aid Effectiveness, RBA seeks to enhance the ownership and responsibilities of partner governments, thus allowing them to decide how they will achieve national goals and objectives. By providing discretion on how outcomes are achieved, RBA seeks to encourage governments to innovate and develop cost-effective ways of achieving these outcomes.

The United Kingdom’s Department for International Development (DFID) initiated an RBA pilot in collaboration with Ethiopia’s Ministry of Education (MoE) in early 2012. The pilot sought to enhance access to and the quality of lower secondary education, which includes grades 9 and 10, among boys and girls and especially among students in Ethiopia’s four designated emerging regions, which are less developed than the country’s seven non-emerging regions. In particular, the RBA pilot was intended to increase the number of grade 10 students sitting for and passing the Ethiopian General Secondary Education Certificate Examination (EGSECE) - 10<sup>th</sup> Grade National Examination in 2012, 2013, and 2014.

To encourage these increases, DFID offered the Government of Ethiopia (GoE) up to £10 million per year for each of three years for increases in the number of students sitting for and passing the EGSECE. The amounts to be provided per additional sitter and passer are shown in Table E.1, with higher amounts for girls than for boys and for students in the emerging regions. Reward payments would be based on the numbers of additional sitters and passers within each region compared with the number of sitters and passers in each region the previous year, thus using an ‘adjusting’ or rolling baseline<sup>1</sup>. The reward payments for additional sitters would be provided irrespective of their performance on the EGSECE.

Table E.1: Per Student Incentives for Sitting and Passing the EGSECE

	Sitter		Passer	
	Emerging region	Non-emerging region	Emerging region	Non-emerging region
<b>Boys</b>	£75	£50	£75	£50
<b>Girls</b>	£100	£85	£100	£85

If the full incentive of £10 million was earned each year, DFID’s modelling estimated that the following increases would occur – *above and beyond what would have occurred in the absence of the pilot*:

- 129,000 more girls and 55,000 more boys would sit for the EGSECE in the non-emerging regions;
- 100,000 more girls and 70,000 more boys would pass the examination in these regions;
- 3,500 more girls and 3,200 more boys would sit for the examination in the emerging regions; and,
- 2,600 more girls and 4,500 more boys would pass the examination in these regions.

<sup>1</sup> In reality there were 44 baselines: girl sitters, girl passers, boy sitters, and boy passers for each of the 11 regions. Reward payments would be provided for increases in sitters or passers for any of these baseline groups.



Several possible explanations exist for increases in the number of sitters and passers independent of the RBA pilot such as trends in the number of students enrolled in Ethiopia's lower secondary schools. Accordingly, a primary purpose of the evaluation was to assess the extent to which changes in the number of sitters and passers could reasonably be attributed to the RBA pilot as opposed to potential competing explanations. The evaluation also addressed several related questions. Did the pilot reduce long-standing inequities between boys and girls and between emerging and non-emerging regions? What are the institutional consequences of the pilot and how have the reward payments been used? What is the optimal level of reward payments for sitters and passers? Did the pilot provide value for money?

To answer these and other questions, DFID awarded a multi-year contract to Mott MacDonald Limited, acting through Cambridge Education, to provide an independent evaluation of the RBA pilot<sup>2</sup>. As part of the evaluation process Mott MacDonald produced an *Inception Report*, a *Baseline Report*, and an interim evaluation report covering the pilot's first two years. The present report provides the final results of the evaluation and addresses seven core issues that DFID identified in its terms of reference for the evaluation. These issues provide the framework for discussion of the results in the chapters that follow; the final chapter provides conclusions and recommendations about the design and implementation of results-based approaches in the education sector.

The evaluation report, intended to be a 'learning document', seeks to assess whether RBA is an effective use of aid to achieve development goals and to inform DFID's decision-making about whether RBA merits use elsewhere. DFID also anticipates that the evaluation will be of value to other donors that are considering the use of RBA in education as well as in other sectors.

The evaluation used a mixed-method approach, relying on both quantitative and qualitative methods, to gather and analyse the data collected. DFID's terms of reference for the evaluation called for an impact evaluation. Such evaluations ideally use evaluation designs that compare a group that receives an intervention with another group that does not receive the intervention, thus allowing comparison of the results achieved by both groups. Responding to the GoE's preference, the RBA pilot was implemented throughout Ethiopia, thus limiting opportunities for valid comparisons. Given this limitation, the evaluation relied on econometric modelling and comparisons of the numbers of sitters during the pilot with the pre-pilot trend and the trend in non-government schools to assess the pilot's relative effects. In addition, the evaluation was based on review of documents, interviews with national, regional and school-level education officials, and attendance at three of the MoE's annual education conferences.

By the end of the three-year pilot, based on increases in the number of sitters and passers of the EGSECE, DFID had provided the GoE with total reward payments of approximately £15.6 million<sup>3</sup>. As noted above, however, the task of the evaluation was to determine whether increases in sitters and passers were attributable to the pilot or to other, alternative explanations. *The alternative explanations prevailed. The evaluation was unable to detect evidence that the RBA pilot improved educational performance for either boys or girls in either the emerging regions or the non-emerging regions.* Although there were changes in the numbers of sitters and passers during the pilot, none of the estimated impacts on the numbers of either boys or girls sitting the EGSECE were either statistically significant or reasonably attributable to the RBA pilot.

---

<sup>2</sup> Although the pilot was initially planned to be for three years, DFID extended the pilot through to December 2015, thus including a fourth round of the EGSECE. Despite this change, this evaluation report covers only three cycles of the EGSECE.

<sup>3</sup> DFID provided approximately £0.9 million, £5.66 million, and £9.0 million in reward payments for results achieved in 2012, 2013, and 2014, respectively.

A primary reason for an absence of any detectable effect is not difficult to discern. The RBA pilot was not well communicated to the regions in time to appreciably affect students' performance. When the evaluation team visited regions during the pilot's first two years, few of their education officials, including head teachers were aware of the pilot. Additionally, the EGSECE's grading system relies on a norm-referenced examination, which provides information on how well a student performs relative to all other students. Such examinations virtually guarantee that some students will not pass the examination regardless of their level of proficiency. Consequently, there is no reason to expect the pilot to have had any effect on the number of students passing the EGSECE other than through its effect on the numbers sitting for it.

In addition to considering whether the pilot had an effect on the number of sitters and passers, the evaluation also assessed whether the pilot successfully reduced inequities related to gender or those between emerging and non-emerging regions. DFID had anticipated extra progress for girls compared to boys, and this was reflected in the higher reward payments offered for girls than for boys. During the pilot some of the gender inequities were reduced, but the improvement did not meet the results associated with maximum fund disbursement. Most important, the evaluation did not find evidence that the RBA pilot accelerated progress toward gender equity. The increases that occurred were largely a continuation of pre-existing trends rather than a change that can be attributed to the RBA pilot.

Similar findings apply to regional inequities. Slightly more than 10 per cent of all 15 and 16 year olds in Ethiopia are estimated to live in the four emerging regions. Accordingly, regional parity would be achieved when a similar percentage of all sitters and passers come from these four regions. The evaluation found no evidence of statistically significant deviations from baseline trends in the numbers of students sitting or passing the examination in either the emerging or non-emerging regions. While there were some gains in the emerging regions during the pilot's three years, the gains were not due to the pilot. Moreover, the gender gap appears to have closed more slowly in emerging regions than elsewhere, and female EGSECE sitters continue to be considerably less likely than males to receive high grades.

Advocates of RBA believe that reliance on country-based systems should not only strengthen these systems but also create incentives to improve them. In addition, according to DFID, reliance on RBA would similarly strengthen the GoE's accountability to its citizens and improve public financial management in the education sector. Given the findings noted above, one would not expect to see major or perhaps even discernible effects on these systems. While there may have been change at the margins, the evaluation was not able to identify any meaningful, pilot-related changes in: (a) the GoE's allocation of resources to lower secondary education; (b) the administration of the EGSECE; (c) the MoE's education management information system; (d) financial management; (e) accountability; or, (f) levels of corruption.

DFID hypothesised that the RBA pilot, with its 'hands-off' approach to development would both change and improve the department's relationship with the GoE. DFID's aspiration that RBA might improve the dialogue between DFID and the MoE has not yet been realised. Although the MoE initially expressed enthusiasm, this positive sentiment was offset by concerns about the nature of the pilot itself. These concerns include a perception of high transaction costs, the relatively small amounts of money available through the pilot, and the absence of resources to pre-fund activities that might increase the numbers of sitters and passers. These amounts pale in comparison with the several hundred million dollars available through traditional and much larger aid projects such as Ethiopia's General Education Quality Improvement Programme, a multi-donor initiative. Interviews with DFID officials likewise reveal recognition that there has not been the anticipated behaviour change on the part of the MoE about RBA.

In the absence of any statistically significant effects of the pilot on the number of passers and sitters it is not possible to conclude that the RBA instrument has offered value for money (VfM) to date. Nonetheless, the analysis does suggest that RBA as an instrument is a low-cost, relatively low-risk approach in the context of support to the Ethiopian education system. Its costs are low if it does not work (it has modest transactions costs but no reward payments) but with potentially high rewards albeit with higher costs if RBA does work.

As noted above DFID used a rolling baseline rather than an estimated counterfactual to calculate reward payments. As a result, DFID risked paying for the results that would have been achieved anyway in the absence of the pilot. For this reason the evaluation team estimated the reward payments associated with baselines other than the one chosen. It did this using the counterfactual estimated as part of the design process and a range of alternative baselines. In each of these alternative baselines DFID would have provided less than the £15.6 million it actually did with the amounts ranging from £2 million to £11 million depending on the alternate baseline chosen. The reward payments would also have been distributed over a different time frame.

Notwithstanding these findings, it is important to emphasise that there are no indications that the reward payments have been wasted or misused. Ethiopia's education system received over £15 million that it would not have received in the absence of the pilot. In a resource-deficient educational system, such as in Ethiopia, any additional resources are assuredly welcome, especially at the school level.

New approaches to aid bring with them the possibility of unintended or unforeseen consequences. Such is the case with the RBA pilot. The evaluation team identified several possible unintended consequences, including: (a) more cheating to increase the number of sitters and passers; (b) pass rates increasing due to changes in the EGSECE; (c) schools focusing attention on students near the EGSECE pass/fail threshold at the expense of others; and, (d) increasing pressures for corruption. Despite the multiple opportunities for unintended consequences, no such occurrences were identified.

In sum, the pilot has been much less successful than its advocates had hoped. The most obvious explanations focus on the pilot's slow start, the tardiness in publicising the pilot and allocating rewards to the regions, the meagreness of the rewards relative to other assured funding from other donor-funded programmes, and perhaps the short duration of the pilot itself. Innovative programmes are rarely adopted quickly, especially when they require considerable changes in long-standing and well-entrenched policies, procedures, and practices. A definitive finger can also be pointed at the use of the rolling baseline. In the pilot's first year the total number of sitters and passers actually declined nationally but increased within some regions. This situation required DFID to provide rewards, albeit small, to the GoE. In the second and third years the number of additional sitters and passers increased, but DFID found itself having to pay for increases that would likely have occurred without the RBA pilot. Knowledge of the size of the cohorts based on enrolments up to grade 9 and their historic transition rates to grade 10 rendered these increases largely predictable.

## Summary of Recommendations

1. If the pilot were to be continued to year 5, DFID and GoE should increase the premiums for girls and the emerging regions based on the analysis of the challenges they continue to face and the use made of the RBA rewards to date.

2. Recognise that, in switching to an RBA approach, expectations of speedy changes in donor and recipient behaviour must be set, on both sides, against the background of many years of working in completely different and often incompatible ways.
3. Recognise that vigorous local ownership and engagement with a new approach such as RBA are a prerequisite and ensure that awareness, buy in, and ownership exist among all levels of stakeholders before implementation begins.
4. When working in a new context, recognise that transaction costs for verification, communication, monitoring, and evaluation are likely to be high. RBA does not create capacity to administer a reward-based system, but such capacity is a prerequisite for success.
5. Reward payments should be based on increases above the trend over some previous period (perhaps five years) rather than the change from the previous year.
6. Financial and capacity-building pump-priming should be considered to allow the initial progress to be made and rewarded.
7. Reward payments should not be based on the number of passers of norm-referenced examinations like the EGSECE.
8. In relation to VfM, recognise that the context of RBA is new, with no right and wrong ways of thinking about VfM, only different interpretations. If DFID wishes to compare the cost effectiveness of its interventions it should ensure the use of common measures and methodologies across the interventions.

Each of these recommendations is discussed in more detail in Chapter 9.

# 1 Results-Based Aid in Ethiopia's Education Sector

This report presents the results of an independent, three-year evaluation of a pilot project intended to improve access to and the quality of lower secondary education in Ethiopia through the use of results-based aid (RBA), an innovative approach to development. After this introductory chapter, which explains the assumptions of the approach as well as its design and implementation in Ethiopia, the following seven chapters discuss the results achieved and the consequences of the pilot observed to date.

## 1.1 Results-based Aid: Theory

RBA is an aid partnership between a donor and a recipient government in which the disbursement of aid is tied to results achieved rather than activities completed or outputs produced. With RBA, donors may or may not pay for inputs or pre-fund projects as is typically the case with donor-funded initiatives. Donors take a 'hands-off' approach and do not direct or specify how the desired results should be achieved. Those decisions are left to the recipient. RBA as designed and delivered in this programme has the following characteristics:

- recipients may need to finance up-front investments in the activities and interventions expected to lead to the desired results;
- *after* previously agreed and measurable outcomes have been achieved and independently verified, a donor disburses resources according to a specified formula;
- the recipient has full responsibility for, and discretion in, deciding how these resources will be used; and,
- the resources are intended to complement other development assistance or domestic resources<sup>4</sup>.

RBA is based on the assumption that financial incentives encourage their potential recipients to increase the delivery of development results. Consistent with the Paris Declaration on Aid Effectiveness, RBA seeks to enhance the ownership and responsibilities of partner governments, thus allowing them to decide how they will achieve national goals and objectives. By providing discretion on how outcomes are achieved, RBA attempts to encourage governments to innovate and develop cost-effective ways of achieving these outcomes. Moreover, RBA does not place any additional financial reporting demands on the recipient.

## 1.2 The RBA Pilot: Design

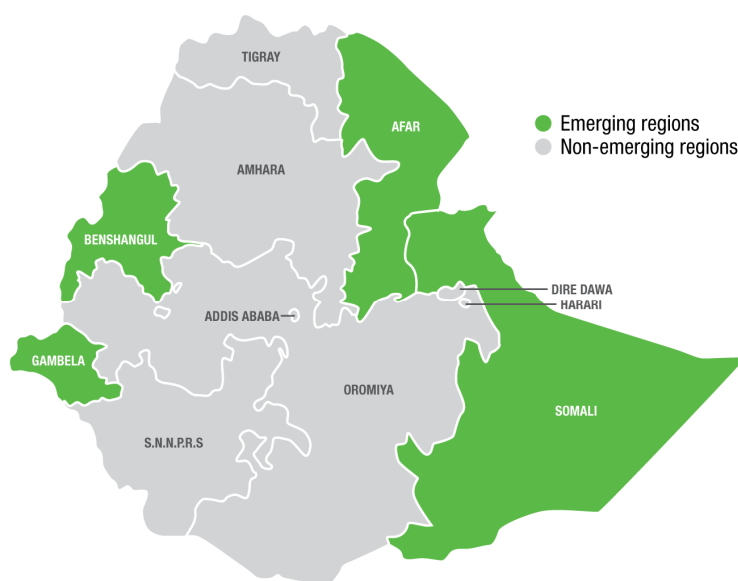
The United Kingdom's Department for International Development (DFID) initiated a three-year RBA pilot in collaboration with Ethiopia's Ministry of Education (MoE) in 2012. The pilot sought to enhance access to and the quality of lower secondary education, which includes grades 9 and 10, among boys and girls and especially among students in Ethiopia's developing regional states (DRS), commonly referred to as emerging regions (see Figure 1.1)<sup>5</sup>. In particular, the RBA pilot was intended to increase the number of grade 10 students sitting for and passing the Ethiopian General Secondary Education Certificate

<sup>4</sup> Center for Global Development, *The Anatomy of Program-for-Results: An Approach to Results-Based Aid*, Working Paper No. 374, 2014. Available at [http://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=2466657](http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2466657)

<sup>5</sup> The four emerging regions include Somali and Afar in the east and Gambella and Benishangul-Gumuz in the west, with pastoral communities in the former and agro-pastoral communities in the latter. As the Government of Ethiopia (GoE) has explained, 'Difficult conditions exist in [these regions], inadequate water in the east leading to nomadic lifestyles and malaria, sleeping sickness and general backwardness in the west. Literacy levels are very low particularly in the pastoral regions and not much different in the agro-pastoral regions as well. The emerging regions are characterised by small, scattered and nomadic populations making it more challenging to provide public services. Most of the areas are inaccessible with poor or no roads and few social services including schools and clinics. There are also very limited personnel in the specialist fields. The Regions also have different ethnic compositions'. See GoE, Ministry of Federal Affairs, *Emerging Regions Development Programme*, 2007. Available at [http://www.uncdf.org/sites/default/files/Documents/erdp\\_54573\\_prodoc\\_0.pdf](http://www.uncdf.org/sites/default/files/Documents/erdp_54573_prodoc_0.pdf)

Examination (EGSECE) - 10<sup>th</sup> Grade National Examination, which is typically administered in May and June of each year. DFID offered the Government of Ethiopia (GoE) up to £10 million per year for each of three years for increases in the number of students (a) sitting for and (b) passing the EGSECE. The reward payments for additional sitters would be provided irrespective of their performance. Encouraging students to sit for and pass the EGSECE clearly addresses an issue of major concern. Recent estimates suggest that in the five years preceding the pilot less than one quarter of girls and one third of boys had completed their lower secondary education and sat for the EGSECE<sup>6</sup>.

Figure 1.1: Ethiopia's Regions



To encourage these increases, DFID agreed to provide the amounts per additional sitter and passer shown in Table 1.1, with higher amounts for girls than for boys and for students in the four emerging regions<sup>7</sup>. The amounts were based on one estimate and two assumptions. The MoE has estimated that the average cost of educating a student for two years of lower secondary education is approximately £50<sup>8</sup>. DFID's *Business Case* assumed that amounts higher than this average cost would encourage the education system to deliver results in excess of what would otherwise have been achieved without the incentive. DFID further assumed the need for still higher incentives for emerging regions (relating to underdevelopment and difficulties of reaching marginalised populations) and girls (socio-economic challenges to attracting and retaining girls in secondary school).

Reward payments would be based on the numbers of additional sitters and passers within each region compared with the numbers of sitters and passers in each region the previous year, thus using an 'adjusting' or rolling baseline<sup>9</sup>.

<sup>6</sup> See Independent Commission on Aid Impact, 'DFID's Education Programmes in Three East African Countries', Report 10, May 2012. Available at <http://www.oecd.org/countries/rwanda/50360183.pdf>

<sup>7</sup> DFID, *Pilot Project of Results-Based Aid (RBA) in the Education Sector in Ethiopia: Business Case*, October 2011, hereafter referred to as the *Business Case*. Available at [iatl.dfid.gov.uk/iati\\_documents/3716785.docx](http://iatl.dfid.gov.uk/iati_documents/3716785.docx).

<sup>8</sup> MoE, *Education Statistics Annual Abstract: 2009-2010*, 2010. Available at <http://www.moe.gov.et/English/Information/Pages/AnnualAbstract.aspx>

<sup>9</sup> In reality there were 44 baselines: girl sitters, girl passers, boy sitters and boy passers for each of the 11 regions. Reward payments would be provided for increases in sitters or passers for any of these baseline groups.

Table 1.1: Incentives for Sitting and Passing the EGSECE

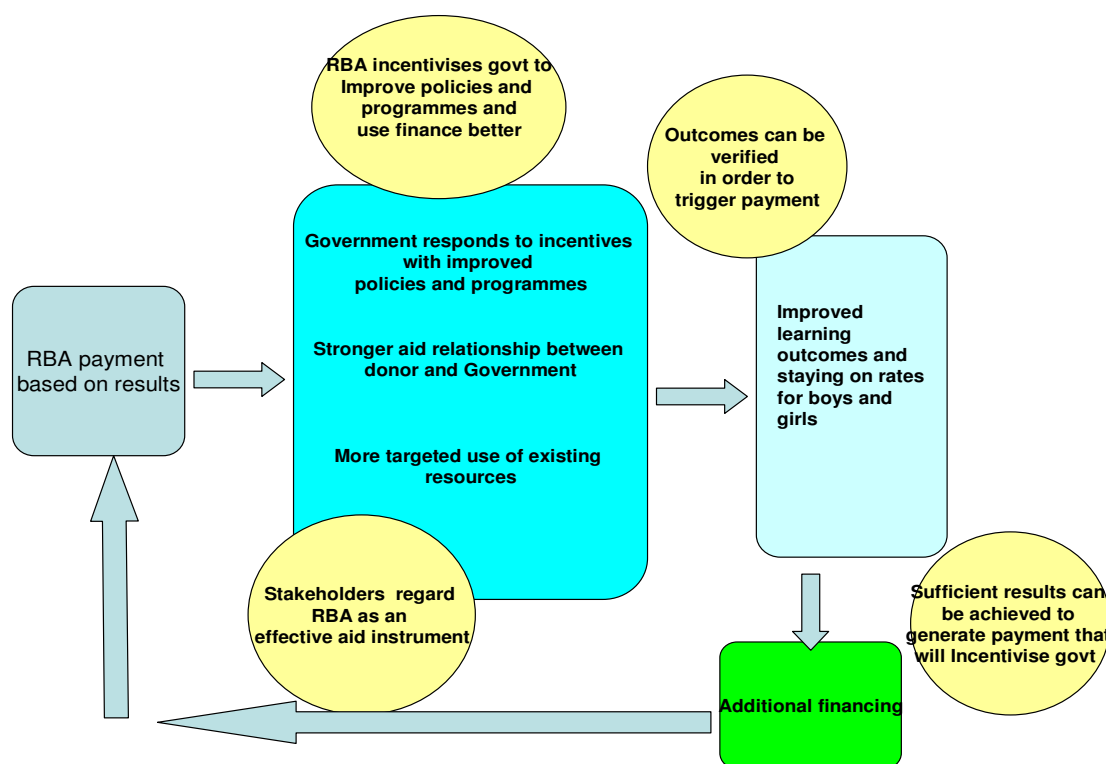
	Sitters		Passers	
	Emerging region	Non-emerging region	Emerging region	Non-emerging region
<b>Boys</b>	£75	£50	£75	£50
<b>Girls</b>	£100	£85	£100	£85

Source: DFID, Business Case.

As shown in Figure 1.2, DFID’s theory of change for the pilot assumed that:

- a relatively small amount of additional resources given to the MoE would incentivise the relevant parts and levels of government to improve their education policies and programmes;
- sufficient results could be achieved to generate reward payments that would incentivise the government;
- outcomes could be verified to trigger disbursement of the financial incentives; and,
- stakeholders would regard RBA as an effective aid instrument.

Figure 1.2: DFID’s Theory of Change for the RBA Pilot in Ethiopia



Source: DFID, Business Case.

The *Business Case* declared that the RBA approach to aid is ‘untested, especially with respect to the hypothesised causal chain between RBA incentives, improved aid relationships, more effective programming and improved results’. For this reason DFID stressed the need for the evaluation to analyse how the theory of change has unfolded by assessing how relations between the incentive scheme, the GoE, and DFID have evolved over the life of the pilot and how those relations have affected the results. These relations have evolved but not necessarily in ways that DFID had anticipated. As a consequence, the evaluation team has refined and revised the theory of change (see Appendix 1).



DFID's *Business Case* further projected a series of expected outputs, outcomes, and impacts (see Table 1.2). As the *Business Case* explained, the 'projections are...not real targets in the conventional sense of the term...[but] are results that are expected to be incentivised by the project'<sup>10</sup>. In other words the annual reward payments that DFID offered to the MoE represented a ceiling on the department's assistance rather than defining or specifying targets that would represent measures of the pilot's success.

Table 1.2: Expected Results of DFID's RBA Pilot in Ethiopia

Expected Results of DFID's RBA Pilot in Ethiopia	
<b>Impacts</b>	Improved access to, and quality of, lower secondary schooling, to be measured by the grades 9 and 10 gross enrolment rate and the percentage of girls and boys passing the grade 10 examination.
<b>Outcomes</b>	An increase in students sitting and passing the EGSECE over an adjusting baseline, especially for girls and in DRS. DFID's modelling predicted the following changes: <ul style="list-style-type: none"> <li>▪ 129,000 more girls and 55,000 more boys sitting the grade 10 examination in non-DRS</li> <li>▪ 100,000 more girls and 70,000 more boys passing the grade 10 examination in non-DRS</li> <li>▪ 3,500 more girls and 3,200 more boys sitting the examination in DRS</li> <li>▪ 2,600 more girls and 4,500 more boys passing the examination in DRS</li> </ul>
<b>Outputs</b>	<ul style="list-style-type: none"> <li>▪ Government responds to incentives with improved policies and programmes that lead to increased enrolment and retention of students in lower secondary school, as measured by the percentage of regional governments responding to RBA incentives with new policies/programmes</li> <li>▪ Incentives lead to more targeted and efficient use of existing resources, as measured by the proportion of sector financing allocated to secondary schooling annually</li> <li>▪ Stronger aid relationship between donor and governments, as measured by the percentage of stakeholders in Ethiopia who perceive RBA to have fewer conditions than other forms of aid in the sector</li> </ul>

Note: DFID's modelling assumed increases of 8 and 7.5 per cent in sitting rates for girls in emerging and non-emerging regions, respectively, over an average increase in the rate without the RBA pilot. For boys, the assumed increase was 5 and 4 per cent, respectively, predicting that an additional 191,000 sitters would be incentivised. For passers, the modelling assumed increases of 80 and 75 per cent in passing rate for girls in emerging and non-emerging regions, respectively, over an average increase in the rate without the pilot. For boys, the assumed increases were 50 and 40 per cent, respectively. The model thus predicted that the pilot would incentivise an additional 178,000 passers. In contrast to these estimated increases, DFID's VfM analysis used slightly different estimated increases in the numbers of sitters and passers, as discussed in Chapter 7.

Source: DFID, *Business Case*.

### 1.3 The RBA Pilot: Action in Response

An initial step for the pilot was the signing of a memorandum of understanding (MoU) between DFID and the GoE in early 2012. The MoU required the GoE to use the payments to reward the achievement of increased numbers of sitters and passers of the EGSECE<sup>11</sup>. The MoE was further required to announce each year's results and allocation of reward payments at the ministry's annual education conference, which typically occurs in October. Other than these requirements, the MoE could disburse and apply the reward payments in any way it wanted.

Best practice in RBA suggests that independent verification of reported results be used to prevent or reduce the incentives for misreporting. Accordingly, to avoid providing rewards for results that did not occur or that were incorrectly or inaccurately reported, DFID required that the number of sitters and passers in the EGSECE be verified independently each year. DFID contracted with Coffey International Development to do so. Coffey's annual verification process involved checking the marking and scoring systems and

<sup>10</sup> DFID's 2013 annual review of the RBA pilot questioned whether it is appropriate to refer to the expected results as 'targets'. See DFID, 'Annual Review: Pilot Project of Results Based Aid (RBA) in the Education Sector in Ethiopia', December 2013. Available at [iati.dfid.gov.uk/iati\\_documents/4341561.docx](http://iati.dfid.gov.uk/iati_documents/4341561.docx).

<sup>11</sup> The MoU also obligated the GoE's adherence to the partnership commitments governing all of the United Kingdom's development assistance. As an example, the MoU required the GoE to respect human rights and to promote transparency and accountability to its citizens.



visiting a sample of schools to ensure that the reported results at the national level matched records at the school level.

With DFID's assistance the MoE prepared guidelines that explained how the reward payments would be distributed to and could be used by the regions<sup>12</sup>. Once the examination results were verified, DFID would provide the reward payments to the Ministry of Finance and Economic Development (MoFED), which would then distribute the payments to the regions based on the MoE's directions. MoFED would also be responsible for all of the pilot's financial management.

The guidelines replicated key principles of DFID's *Business Case*, including the reward payment levels shown in Table 1.1. In addition, the guidelines emphasised the regions' freedom to decide how they could use these payments. The primary requirement was that funds be devoted to 'any item that is perceived to improve the access and quality of education, leading to improved performances in the EGSECE'. The guidelines also required each regional education bureau (REB) to develop plans outlining how it would use the region's RBA rewards.

For each of nine subjects in the EGSECE on which students are tested, they are assigned a score ranging from 0.0 to 4.0. To achieve a grade of A or 4.0 for a single subject, for example, a student's raw score (number of questions answered correctly) must be at least two standard deviations above the mean score achieved for all students in that subject. A grade of B or 3.0 would be given to a student whose score in a subject is between one and two standard deviations above the mean<sup>13</sup>. To determine an overall grade point average (GPA), scores in English and mathematics, which are compulsory subjects, plus a student's five highest scores from among the other subjects are added and then divided by seven. That process provides an overall composite score, which also ranges from 0.0 to 4.0. A GPA of 2.0 represents a passing score on the EGSECE.

The EGSECE is thus a norm-referenced assessment. Such assessments indicate how a student has performed relative to other students taking the same test but do not indicate whether a student has achieved a particular level of proficiency or competence in the subjects tested<sup>14</sup>. Ethiopia's National Educational Assessment and Examinations Agency (NEAEA) prepares the EGSECE each year. Although the agency is developing an item bank that would permit re-use of questions and comparison of scores across years, it has not used such a bank in the past. With different questions each year the relative difficulty of the EGSECE also changes each year. This means that results from one year cannot be compared with results from other years. Scores on the examination determine receipt of a grade 10 completion certificate as well as eligibility for advancement to grade 11 and the university track.

Table 1.3 provides a timeline for the project. Table 1.4 shows the relation between the Ethiopian calendar (EC), the Gregorian calendar (GC), and each of the three pilot years. Unless otherwise noted, all subsequent dates for all years use the Ethiopian calendar rather than the Gregorian calendar.

<sup>12</sup> MoE, *Results Based Aid (RBA) Pilot Guidelines: Improving access, quality and equity in general secondary education*, EC 2004, August 2012. As the *Business Case* noted, 'Feedback received from the MoE during the design phase indicates that [it] would allocate the additional funding to regions on the basis of results and regions would allocate to districts and schools also based on results achieved'.

<sup>13</sup> Students with scores within  $\pm 0.99$  standard deviations of the examination's mean score would receive a C or 2 points; those within -1.0 to -2.0 standard deviations of the mean would be given a D or 1 point. An E or no points would be awarded for students scoring two or more standard deviations below the mean examination score.

<sup>14</sup> In contrast, a criterion-referenced examination has a fixed 'passing' score, thus making it possible for all students to pass or fail the examination.

Table 1.3: Timeline for the RBA Pilot in Ethiopia

Pilot year 1	
September 2011	Start of school year
February 2012	DFID and the GoE sign memorandum of understanding on the RBA pilot
May-June 2012	EGSECE administered
August 2012	Results of EGSECE available
August 2012	MoE completes guidelines for administration of RBA pilot and its rewards
Pilot year 2	
September 2012	Start of school year
September 2012	DFID's projected date for distribution of year 1 reward payment to GoE
October 2012	First RBA reward payment of £0.9 million announced at annual education conference
March 2013	DFID provides reward payment to GoE based on year 1 EGSECE results
May-June 2013	EGSECE administered
August 2013	Results of EGSECE available
Pilot year 3	
September 2013	Start of school year
September 2013	DFID and MoE agree that year 1 and year 2 payments can be combined
October 2013	Second RBA reward payment of £5.66 million announced at annual education conference. RBA guidelines from August 2012 distributed to regions, which are asked to prepare plans describing how they will spend their reward payments. Uncertainty exists about whether the plans must be submitted to the MoE.
December 2013	DFID provides reward payment to GoE based on year 2 EGSECE results
April 2014	MoE appoints RBA lead
April 2014	MoE and MoFED reach agreement on how payments will be distributed
May-June 2014	EGSECE administered
May-August 2014	Reward payments for years 1 and 2 distributed to regions
August 2014	Results of EGSECE available
October 2014	Third RBA reward payment of £9.0 million announced at annual education conference.
December 2014	DFID provides reward payment to GoE based on year 3 EGSECE results
Jan.-Feb. 2015	Year 3 reward payments distributed to regions.

Table 1.4: Comparison between Ethiopian, Gregorian, and RBA Pilot Calendars

Year		
Ethiopian calendar*	Gregorian calendar	RBA pilot calendar
2003	2010/2011	Initial baseline year
2004	2011/2012	Year 1
2005	2012/2013	Year 2
2006	2013/2014	Year 3

\* The Ethiopian New Year starts in September of each Gregorian year. Accordingly, EC 2003 began in GC September 2010.

As noted above, the pilot used a rolling baseline. For the pilot's first year (EC 2004 or year 1), the number of sitters and passers in EC 2003 served as the baseline for the results in EC 2004. The results in that year served as the baseline for the results achieved in EC 2005 (year 2) and EC 2005 served as the baseline for the results achieved in EC 2006 (year 3).

## 1.4 The RBA Pilot: What Happened?

Table 1.5 shows the number of sitters and passers from the initial baseline year, EC2003, and the three subsequent years, all of which were part of the pilot. Although the number of sitters and passers increased in several regions between the first baseline year, EC 2003, and EC 2004 (year 1), there was an overall *decrease* in both passers and sitters. Despite the overall decline, DFID’s agreement with the MoE required payment of the incentive based on the increases that occurred within each region – but *not* on the net change at the national level. As an illustration, in Somali the number of sitters increased by 3,249 and the number of passers by 3,186 during the pilot’s first year, so DFID provided an incentive payment for these and all other regional increases that occurred. In the second year (EC 2005) the numbers of sitters was higher than in the previous year but still below the original baseline year (i.e., EC 2003). Nonetheless, because the numbers of sitters and passers in most regions were higher than in the previous year, additional incentive payments were mandated and provided.

Table 1.5: EGSECE Sitters and Passers, EC 2003 to EC 2006

	Sitters				Passers			
	EC 2003	EC 2004	EC 2005	EC 2006	EC 2003	EC 2004	EC 2005	EC 2006
	GC 2010/11	GC 2011/12	GC 2012/13	GC 2013/14	GC 2010/11	GC 2011/12	GC 2012/13	GC 2013/14
Addis Ababa	41,177	38,918	43,382	40,469	33,890	31,953	34,458	31,054
Afar	2,327	2,847	2,435	2,878	1,630	2,098	1,790	1,997
Amhara	129,272	117,626	130,003	164,514	91,527	85,322	96,631	110,579
Benishangul - Gumuz	6,781	6,826	7,034	6,507	4,145	4,039	3,721	3,415
Dire Dawa	3,360	3,562	3,594	3,374	1,951	1,931	2,022	2,390
Gambella	4,096	4,932	6,176	6,531	2,049	2,716	3,130	2,633
Harari	1,989	1,133	1,321	1,644	1,377	928	1,072	1,262
Oromiya	211,325	178,161	184,672	216,437	130,900	112,851	118,762	119,769
SNNPR*	107,558	106,001	106,930	129,850	67,597	66,829	73,800	77,210
Somali	6,883	10,132	13,619	15,591	5,492	8,678	11,837	12,583
Tigray	48,567	49,678	57,615	73,756	34,117	32,142	39,307	39,793
<b>Total</b>	<b>563,335</b>	<b>519,816</b>	<b>556,781</b>	<b>661,551</b>	<b>374,675</b>	<b>349,487</b>	<b>386,530</b>	<b>402,685</b>

\* SNNPR = Southern Nations, Nationalities, and Peoples’ Region.

Note: The totals include regular and evening students only but exclude a small number of expatriate Ethiopian students in schools in Saudi Arabia. After verification of the numbers of sitters and passers shown in the table, DFID used these numbers to determine the size of each year’s reward payments (see Table 1.6, below). Readers should note that the numbers of sitters and passers shown for EC 2005 and EC 2006 do not correspond with the comparable numbers in the MoE’s Education Statistics Annual Abstracts for these years. The numbers in the two abstracts are in conflict. The EC 2005 abstract reported 547,791 sitters and 384,203 passers for the EC 2005 EGSECE. In contrast, for the same EGSECE the EC 2006 abstract reported 756,637 sitters and 463,261 passers. In addition, the numbers of sitters and passers shown in the two abstracts include regular and evening students plus those in private or non-government schools. As discussed below, private and non-government schools were not eligible for reward payments. Source: DFID, ‘Annual Review: Pilot Project of Results-Based Aid (RBA) in the Education Sector in Ethiopia’, December 2014.

Based on the increases in the numbers of sitters and passers over the pilot’s three years, DFID provided a total of £15.6 million (see Table 1.6), or slightly more than half the £30 million that had been allocated. At this point several important issues must be noted. First, DFID was overly optimistic in its expectations for the pilot’s implementation as well as for the time required to achieve the expected results. DFID projected that it would provide the first reward payment to the GoE in September 2012, with the ‘implementation

phase' beginning immediately thereafter<sup>15</sup>. DFID similarly indicated its expectation that the pilot's results could be achieved and assessed no later than early 2014 thus permitting the completion of a draft evaluation report by July 2014.

These expectations did not coincide with reality. Although the MoE had completed its guidelines on the distribution of the reward payments to the regions in August 2012, neither the guidelines nor information about the RBA pilot were shared with the regions until August 2013. This meant that *two cycles of the EGSECE within the pilot period had already been completed before the guidelines were distributed to the regions*. Furthermore, few head teachers were aware of the RBA pilot during its first two years<sup>16</sup>.

Table 1.6: RBA Rewards (£) Allocated to Regions

Region	EC 2004	EC 2005	EC 2006	Total
	Year 1	Year 2	Year 3	Years 1-3
Addis Ababa	36,972	469,578	293,022	799,572
Afar	45,006	15,563	48,465	109,034
Amhara	101,937	1,521,551	2,672,672	4,296,160
Benishangul-Gumuz	11,117	60,723	51,359	123,199
Dire Dawa	9,556	21,143	34,860	65,559
Gambella	70,914	105,286	58,637	234,837
Harari	1,017	19,165	26,563	46,745
Oromiya	143,343	1,401,824	2,408,135	3,953,302
SNNPR	84,059	835,315	1,687,299	2,606,673
Somali	279,066	362,542	248,295	889,903
Tigray	104,310	795,597	1,016,147	1,916,054
Total to regions	887,297	5,608,286	8,545,454	15,041,038
Retained by the MoE	8,963	56,649	449,761	515,373
<b>Total</b>	<b>896,260</b>	<b>5,664,935</b>	<b>8,995,215</b>	<b>15,556,411</b>

Second, DFID committed to provide the amounts shown in Table 1.1 per additional EGSECE sitter and passer. In Ethiopia most students in grade 10 attend public, government-supported schools but about 5 per cent of all sitters (and about 8 per cent of all passers) each year attend non-government schools. The MoE's guidelines were clear that non-government schools would not be eligible to receive any reward payments. The consequence was that the RBA pilot did not incentivise any of the non-government schools to do anything to improve access or educational quality<sup>17</sup>.

Third, due to the small amount of the incentive payment received for the increases that had occurred in the pilot's first year, the MoE chose to delay until May, June, and July 2014 the distribution to the regions of the funds that had been earned in the pilot's first two years. To the extent that schools within these regions received any funds, the funds were distributed after the EGSECE's third cycle during the pilot (i.e., in EC 2006). Although the MoE's guidelines for the RBA pilot specified that each region's allocation would be 'based solely on regional improvements in [grade 10] sitters and passers', the MoE chose not to use this formulation – and it did so without revising the guidelines. To ensure that every region received some of

<sup>15</sup> DFID, Terms of References (TOR) 'Evaluation of Pilot Project of Results-Based Aid (RBA) in the Education Sector – Ethiopia', 2011.

<sup>16</sup> Coffey International, 'Phase 4 Verification Report, Independent Verification of Key Government of Ethiopia (GoE) Educational Data for a Pilot of Results-Based Aid (RBA)', January 2015.

<sup>17</sup> Other groups of sitters and passers also exist and are excluded from all subsequent tables, figures, and analyses. There are a small number of sitters and passers who had failed the EGSECE in previous years. They are typically not assigned to a particular school.

the RBA rewards, the MoE decided that half of the rewards would be allocated on the basis of the number of additional sitters and passers per region and half on the basis of the total number of grade 10 sitters and passers per region<sup>18</sup>. The MoE also decided that it would retain only 1 per cent of reward payments for EC 2004 and EC 2005 (and 5 per cent in EC 2006) rather than the 7 per cent specified in the guidelines for contingency, monitoring, co-ordination, and evaluation.

If allocations had been based solely on the increases in the number of sitters and passers, not all regions would have received a portion of the first year's incentive payment. In contrast, the MoE's revised formula ensured that every region received a portion of all payments. In some regions the amounts received were considerably more than if the payments had been based solely on the increase in number of sitters and passers. Benishangul-Gumuz did not earn any rewards in EC 2006; the numbers of its sitters and passers declined from EC 2005 to EC 2006 but the region still received over £50,000. Other regions received less than what they would have received according to the original guidelines. As an example, using the reward amounts shown in Table 1.1, Somali would have received almost £1.1 million for its improved performance in EC 2004 and EC 2005. Using the MoE's revised guidelines, the region received about £640,000. Appendix 2 provides other illustrations of how the actual allocations differed from the allocations that would have been provided had the MoE adhered to its original (and unrevised) guidelines.

## 1.5 The RBA Pilot: The Evaluation

Although the data in Tables 1.5 and 1.6 are of interest, they do not address several key questions. To what extent are the changes in the number of sitters and passers attributable to the RBA pilot or to other, concurrent initiatives in the education sector in Ethiopia? Did the pilot reduce long-standing inequities between boys and girls and between emerging and non-emerging regions? Several possible explanations exist for changes in the number of sitters and passers. For example, between EC 2003 and EC 2006 the number of secondary schools in Ethiopia increased by more than 50 per cent, and there was a 17 per cent increase in the number of girls enrolled in grades 9 and 10<sup>19</sup>.

Other questions are no less important. What are the institutional consequences of the pilot and how have the reward payments been used? Did the pilot provide value for money? Were the rewards set at the optimal level? DFID noted its particular interest in 'establishing rigorously that the results achieved as a consequence of the RBA pilot were greater than would have been achieved without the RBA pilot'<sup>20</sup>.

To answer these questions, DFID awarded a multi-year contract to Mott MacDonald, which Cambridge Education implemented, to provide an independent evaluation of the RBA pilot<sup>21</sup>. As part of the evaluation process Mott MacDonald produced an *Inception Report*, a *Baseline Report*, and an interim evaluation report covering the pilot's first two years<sup>22</sup>. The present report provides the final results of the evaluation and addresses seven core issues (see Table 1.7) that DFID identified in its terms of reference (ToR; see

<sup>18</sup> DFID, 'Annual Review: Pilot Project of Results-Based Aid (RBA) in the Education Sector in Ethiopia', December 2013. Available at [iati.dfid.gov.uk/iati\\_documents/4341561.docx](http://iati.dfid.gov.uk/iati_documents/4341561.docx). According to the 2014 Annual Review, 'The MoE wished to provide an incentive for regions to improve, but felt that disbursing funds only based on performance would restrict the ability of poor-performing regions to improve'. Available at [iati.dfid.gov.uk/iati\\_documents/4839826.docx](http://iati.dfid.gov.uk/iati_documents/4839826.docx).

<sup>19</sup> MoE, *Education Statistics Annual Abstract, EC 2006*.

<sup>20</sup> DFID, Terms of References (ToR) 'Evaluation of Pilot Project of Results-Based Aid (RBA) in the Education Sector – Ethiopia', 2011.

<sup>21</sup> Although the pilot was initially planned to be for three years, DFID extended the pilot through to December 2015, thus including a fourth round of the EGSECE. Despite this change, this evaluation report covers only three cycles of the EGSECE (i.e., EC 2004 through to EC 2006).

<sup>22</sup> Terry Allsop, et al., *Evaluation of the Pilot Project on Results-Based Aid in the Education Sector in Ethiopia: Inception Report* (Cambridge: Cambridge Education), June 2012; Terry Allsop, et al., *Evaluation of the Pilot Project on Results-Based Aid in the Education Sector in Ethiopia: Baseline Report* (Cambridge: Cambridge Education), July 2013; Terry Allsop, et al., *Evaluation of the Pilot Project on Results-Based Aid in the Education Sector in Ethiopia: Year 1 and Year 2 Report, EC 2004-EC 2005* (Cambridge: Cambridge Education), May 2014.

Appendix 3) for the evaluation. These issues provide the framework for discussion of the results in the following seven chapters; the final chapter provides conclusions and recommendations about the design and implementation of results-based approaches in the education sector.

Table 1.7: Evaluation Questions Addressed in This Report

Chapter	Issue
2	Results/Value added: To what extent did the RBA pilot increase educational results (compared with other/traditional methods including Ethiopia's General Education Quality Improvement Programme and Promoting Basic Services)? Are these changes attributable to RBA?
3	Equity: Who benefited from these improved results? Have disparities (boys versus girls, emerging regions versus others) declined?
4	System effects: Effects on resource allocation, education information systems, accountability, financial management, corruption).
5	Understanding why the approach works (or does not work): What factors and processes have been responsible for the results?
6	Impact on aid relationships: Has the nature of the DFID/GoE dialogue improved?
7	Value for money: Do the benefits of the programme outweigh the costs? Are RBA incentives set at the optimal level?
8	Unintended consequences: To what extent has the programme resulted in unintended consequences? Have these been positive or negative? Were they identified early and remedial actions taken as necessary?

## 1.6 The Evaluation's Purpose, Methods, and Limitations

In commissioning the evaluation, DFID's expectation was that the lessons learned plus those from a related education-based RBA pilot in Rwanda would be of value to other donors and inform the department's decision-making about the merits of extending the approach to other countries and sectors. As DFID explained in its *Business Case*, 'the purpose of this evaluation will be to assess whether RBA is an effective use of foreign aid to achieve development goals' and 'a measure of the pilot's overall success will be the extent to which other stakeholders, including GoE and other donors, perceive RBA as being an effective instrument and whether it leads to similar arrangements either in the education sector or beyond'.

The evaluation team used a mixed-method approach to gather information for the evaluation. The team's *Inception Report* provided extensive discussion of the methods to be used. Chapters 2, 3, and 7 are primarily quantitative and respond to DFID's request to provide a 'rigorous assessment of the results achieved against a credible counterfactual'. Each of these chapters is accompanied by an appendix that explains in detail the assumptions made and the methods used. In choosing and applying these methods, the evaluation team has benefitted from several discussions with Upper Quartile, the organisation that DFID contracted to conduct the evaluation of the RBA education pilot in Rwanda.

Chapters 4, 5, 6, and 8 draw primarily on qualitative research. This research involved an extensive review of project-related materials, attendance at each of the MoE's annual education conferences during the pilot, and interviews with head teachers, teachers, students, education officials at all levels, and members of parent teacher student associations (PTSA) in 30 schools across the eleven regions. These schools were chosen to be representative of their regions and so of the country. In year 1 the research involved interviews with MoE officials. In year 2 interviews were conducted with MoE officials and officials from five REBs. In year 3 repeat interviews were conducted with MoE officials as well as the RBA leads and heads of all eleven REBs. Interviews were also conducted with head teachers, teachers, students and district-level officials from nine schools across six regions where there was some evidence of RBA activity by the end of year 3. They were selected because they illustrated the RBA process at the school level. They were



representative of their regions but not of all Ethiopia. In keeping with the general principles of qualitative research and to encourage candour and openness, respondents were promised anonymity.

Standard qualitative research measures were used to ensure the validity of this qualitative data and address the potential for bias<sup>23</sup>. Semi-structured interview schedules were designed for each cohort (i.e. REB staff, head teachers, etc.) to ensure that key issues were considered across the study while also giving interviewees the opportunity to discuss other issues of importance and relevance to them. The main points from each interview were summarised and agreed between the researchers and interviewees to reduce the risk of bias. Data from each interview were checked for internal consistency (i.e., to ensure that there were no contradictions in the individual accounts). The information collected was then triangulated by comparing the accounts on a region-by-region basis and across the entire country.

The evaluation team faced several limitations in completing this report. On the one hand, not all data were available in time while the quality of other data were occasionally in doubt. As an illustration, problems with data from the emerging regions in the MoE's education management information system (EMIS) imposed some limitations on the analysis in Chapter 3, which focuses on gender and regional equity. Trying to discern the possible unintended effects of the pilot on the country's education system is challenging when other possible explanations exist and some of the consequences are difficult to detect, such as cheating in a school or corruption within a region.

On the other hand, DFID's ToR called for an impact evaluation and required the identification of a 'credible counterfactual'<sup>24</sup>. As the ToR further explained, DFID was also interested in assessing attribution, which implies causality, between the pilot and the results observed and to demonstrate that 'the results achieved as a consequence of the RBA pilot were greater than would have been achieved without the RBA pilot'<sup>25</sup>. Impact evaluations ideally use experimental or quasi-experimental methods. These methods typically compare a group that receives an intervention with another group (the counterfactual) that does not receive the intervention, thus allowing comparison of the results achieved by both groups.

Despite the desirability of these methods, the ToR recognised that they were not feasible. As a consequence the evaluation team has relied on sophisticated econometric modelling to assess whether the results discussed in this report are plausibly linked to the RBA pilot<sup>26</sup>. There are different opinions about the relative strengths and limitations of such modelling, but the evaluation team is confident that its models are methodologically robust and appropriate, as are the assumptions upon which the modelling is based<sup>27</sup>. As the evaluation team noted in its *Inception Report*, 'However one views these assumptions, they are necessary in the absence of a contemporaneous comparison group. In that situation, the only information we have about education outcomes in the absence of RBA must come from the pre-RBA period and must be projected forward into the RBA period'.

<sup>23</sup> Amare Asgedom and Barbara Ridley, 'Historical Narratives in Ethiopia', in Paul Smeyers, et al., ed., *International Handbook of Interpretation in Educational Research*, 2015.

<sup>24</sup> DFID, Terms of References (ToR) 'Evaluation of Pilot Project of Results-Based Aid (RBA) in the Education Sector – Ethiopia', 2011.

<sup>25</sup> According to the Organisation for Economic Co-operation and Development's *Glossary of Key Terms in Evaluation and Results Based Management* (2002), attribution reflects the 'ascription of a causal link between observed (or expected to be observed) changes and a specific intervention'.

<sup>26</sup> DFID's 2014 annual review of the pilot supports the team's use of econometric models: "Measuring the effect of RBA on education outcomes requires an impact evaluation. As the project is nationwide there are no regions or schools that can be used as a control group. This means econometric modelling of a 'counterfactual' – an estimate of how many sitters and passers had occurred in the absence of the RBA is required". Moreover, the annual review also declares that the econometric modelling that the evaluation team has used, interrupted time series (ITS), 'is in fact one of the strongest quantitative impact estimation approaches to measure impact of the pilot'. The annual review is available at [iati.dfid.gov.uk/iati\\_documents/4839826.docx](http://iati.dfid.gov.uk/iati_documents/4839826.docx)

<sup>27</sup> The evaluation team's *Inception Report* contains a lengthy discussion of the strengths and limitations of its proposed approach. See Terry Allsop, et al., *Evaluation of the Pilot Project on Results-Based Aid in the Education Sector in Ethiopia: Inception Report* (Cambridge: Cambridge Education), June 2012.

## 1.7 An Independent Evaluation

DFID's ToR mandated that the evaluation of the RBA pilot be independent, and Mott MacDonald is comfortable in declaring that it is. No member of the evaluation team (or any staff member of Cambridge Education) was involved in the pilot's design or with the GoE's response to the pilot, and none had any real or perceived conflict of interest during the evaluation. Furthermore, Mott MacDonald's corporate ethics policy 'prohibits the offering, giving, solicitation or acceptance of any bribe, whether cash or other inducement, or engaging in any other corrupt practice' and all the company's employees and consultants are 'required to avoid situations that could interfere, or appear to interfere, with the impartial discharge of their duties'. At no time during the evaluation was any team member subjected to outside pressure or efforts – from any person or organisation – to sway or influence the results reported in this evaluation.

Although there were some small differences of opinion within the evaluation team in preparation of the present report these differences were all reconciled before the report's completion. Likewise, although team members were responsible for preparing separate chapters, all team members were asked to review the entire report to ensure their understanding of and concurrence with its content.

Finally, it should be noted that the ToR includes a statement that the evaluation team would be expected to provide 'ongoing advice and guidance' to DFID and the MoE. Despite this expectation, the evaluation team explained to DFID/Ethiopia that providing advice or guidance could compromise the team's and the evaluation's independence. DFID accepted this explanation, and the evaluation team did not provide any advice or guidance to either DFID or the MoE during the evaluation with the exception of a few suggestions that were included in the evaluation team's interim report on the RBA pilot<sup>28</sup>.

---

<sup>28</sup> The evaluation team had also suggested that a different baseline be used, but DFID did not accept this suggestion. Chapter 7, which discuss the pilot's value for money, considers the possible consequences of using alternative or counterfactual baselines.



## 2 The RBA Pilot's Results and Value Added

### 2.1 Impacts on the Numbers of Sitters Attributable to the RBA Pilot

A central question for this evaluation is the estimation of the impact of the RBA pilot on the numbers of sitters and passers. There is no evidence that the RBA pilot improved educational performance for either boys or girls in either the emerging regions or the non-emerging regions. This chapter discusses these findings.

Impact estimation is always difficult. It requires an assessment of what would have happened in the absence of the intervention (the 'counterfactual'), which cannot be observed directly. Impact estimation is especially difficult in the case of the RBA pilot because the approach was initiated and implemented simultaneously and nationwide, leaving no group of government schools within the country unexposed to the pilot that might be used as a comparison group to represent the counterfactual. In the absence of a valid counterfactual, it is not possible to employ experimental or quasi-experimental evaluation designs that are typically the most rigorous and the most desirable means of assessing the impacts of interventions such as the RBA pilot.

The estimation model proposed in the evaluation team's *Inception* and *Baseline* reports to address this issue is an Interrupted Time Series (ITS) or 'deviation from trend' model, estimated at the school level. The ITS model uses a trend line fitted to observations in the baseline period to predict the values of the desired outcomes that would have occurred in the absence of the RBA pilot. In the current analysis, a comparative interrupted time series (CITS) model is used<sup>29</sup>. The CITS model estimates the pilot's impacts as deviations from the baseline trends in numbers of sitters and passers in government schools, adjusted for changes in the same outcomes in non-government schools. The model improves on the ITS model in several ways (for details, see Appendix 4):

- It uses a curvilinear trend, rather than a simple linear trend, to conform better to the baseline trend;
- It includes a comparison group of non-government schools, which explains over 80 per cent of the variation in the numbers of students sitting for and passing the EGSECE in the baseline period;
- It incorporates controls for variation in student cohort size over time;
- It is estimated at the zonal level, rather than the school level, to capture the effects of changes in the number of schools; and,
- It allows each zone to have its own level of trend line, to reflect variation in performance across the sample<sup>30</sup>.

In a situation in which an intervention is implemented simultaneously and nationwide, as was the RBA pilot in Ethiopia, the CITS model is the strongest evaluation model available, as witnessed by its use in the evaluation of the 'No Child Left Behind' education reform legislation in the United States by two teams of top-rank researchers<sup>31</sup>.

<sup>29</sup> In developing this model the evaluation team reviewed the models used in the DFID-sponsored RBA pilot in Rwanda and adopted those features that are relevant to the Ethiopian pilot and applicable with the data available in Ethiopia. See Upper Quartile, 'Evaluation of Results-Based Aid in Rwandan Education: Econometric Report 2014 (draft)', October 2014.

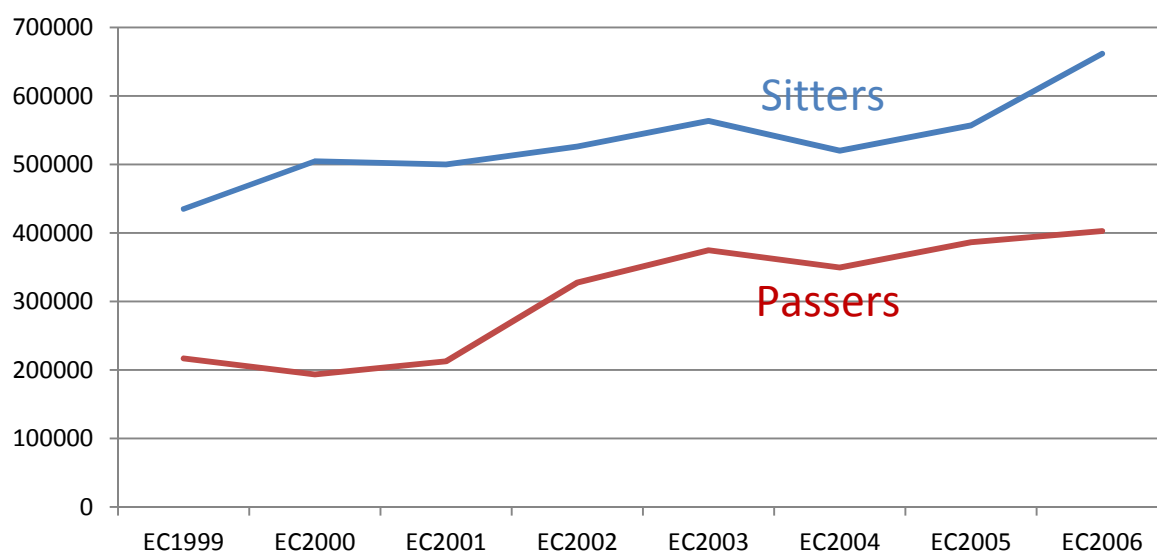
<sup>30</sup> A zone is an administrative subdivision of a region.

<sup>31</sup> Thomas Dee and Brian Jacob, 'The Impact of No Child Left Behind on Student Achievement', *Journal of Policy Analysis and Management*, Summer 2011, and Manyee Wong, Thomas D. Cook, and Peter M. Steiner, *No Child Left Behind: An Interim Evaluation of Its Effects on Learning Using Two Interrupted Time Series Each With Its Own Non-Equivalent Comparison Series*, Institute for Policy Research, Northwestern University, Working Paper #WP-09-11, 2009.

The numbers of EGSECE sitters and passers are measured by the NEAEA, which administers the examination. Regional totals are published each year in the MoE's *Education Statistics Annual Abstract*. The NEAEA provided the evaluation team with school-level data on the numbers of sitters and passers. Coffey International, the company responsible for verifying these data, has confirmed the data to be reliable.

The aggregate numbers of grade 10 students sitting and passing the EGSECE fell in EC 2004, the first year of the RBA pilot, and rose again in EC 2005 and EC 2006, RBA years 2 and 3 (see Figure 2.1). A major portion of the reduction in numbers of students sitting and passing the EGSECE in EC 2004 can be attributed to a drop in grade 10 enrolment compared with the previous academic year. Similarly, the increase in the numbers of sitters and passers from EC 2004 to EC 2006 reflects a rebound in enrolment from the temporarily low levels of EC 2004. Therefore, in estimating the impacts of RBA during the pilot period the analysis controls for cohort size.

Figure 2.1: Numbers of Grade 10 Students Sitting and Passing the EGSECE, EC 1999-2006



Source: NEAEA.

As this discussion makes clear, although RBA funds were provided on the basis of observed changes in the numbers of students sitting and passing the EGSECE, those changes may not have been entirely attributable to the RBA pilot. Accordingly, the CITS model was used to separate the changes attributable to RBA from those due to other factors. Appendix 4 provides a technical discussion of the models used in this chapter, the specification tests used to determine the best model to estimate the pilot's impact, and the precision of the estimates.

The resulting estimates for EC 2004-2006 (RBA years 1-3) are shown in Table 2.1. The estimates are expressed as percentage differences from the value that would have existed in the absence of the RBA pilot if past trends in the numbers of sitters and passers had continued. For example, the estimate for boys in emerging regions in year 1 (-1.9 per cent, the top left number in the table) indicates that the number of students in this group sitting for the exam was 1.9 per cent below the value that would have been expected on the basis of the projected trend. A t-statistic, a measure of statistical uncertainty, is shown in parentheses below each impact estimate. Estimates are usually deemed to be 'statistically different from

zero', and therefore representing real non-zero effects, only if their t-statistic exceeds 1.96<sup>32</sup>. The t-statistic for boys in emerging regions in year 1 is -.21 (far less than the 'required' minimum value of 1.96), and this indicates that the estimate for these boys is not statistically different from zero and thus does not provide evidence of a real non-zero effect.

It is important to recognize that these estimates have a range of uncertainty attached to them – the so-called 'confidence interval' or 'margin of error'<sup>33</sup>. As a measure of this uncertainty, Table 2.1 and subsequent tables in this chapter show the upper bound of the 95 per cent confidence interval – the range at which one can be 95 per cent confident includes the true RBA impact (see discussion in Appendix 4). This value is shown in square brackets below the t-statistic.

Table 2.1: Impacts, Years 1-3, Numbers of Boys and Girls Sitting for the EGSECE

	Boys			Girls		
	Year 1	Year 2	Year 3	Year 1	Year 2	Year 3
<b>Emerging regions</b>	-1.9% (-.21)	4.6% (.43)	7.7% (.63)	-11.3% (-1.04)	-4.6% (-.36)	-4.1% (-.28)
	[12.9%]	[22.1%]	[27.7%]	[6.5%]	[16.4%]	[19.9%]
<b>Non-emerging regions</b>	9.3% (1.41)	14.2% (1.85)	13.2% (1.49)	4.8% (.65)	-1.0% (-.11)	-3.4% (-.34)
	[20.1%]	[26.8%]	[27.7%]	[16.9%]	[13.9%]	[13.0%]
<b>National</b>	5.9% (.86)	14.2% (1.78)	12.9% (1.40)	-1.5% (-.20)	-3.5% (-.39)	-7.3% (-.70)
	[17.2%]	[27.3%]	[28.0%]	[10.8%]	[11.2%]	[9.8%]

Note: The t-statistics for each estimate are in parentheses. The upper bound of the 95 confidence interval (one-tailed test) is shown in square brackets below each t-statistic.

Source: Regression analysis of data from the MoE and the NEAEA, EC 1999-2006.

None of the estimated impacts on the numbers of either boys or girls sitting the EGSECE in the first three years of RBA were statistically significant<sup>34</sup>. Accordingly, *these estimates provide no evidence that RBA increased the number of students sitting for the examination during these years, either nationally or in the emerging or non-emerging regions taken separately*. The upper bounds of the 95 per cent confidence intervals suggest that the true impact could have been as great as 28 per cent for boys in year 3, in both the emerging and non-emerging regions, and as large 20 per cent for girls in the emerging regions and 13 per cent for girls in the non-emerging regions. By definition, however, there is only a 5 per cent chance that the true impacts were this large – compared with a 95 per cent chance that they were not.

To test the robustness of the estimation model, the analysis also estimated impacts on the number of sitters with two alternative models. One, suggested by DFID, used the CITS model to estimate impacts in year 3, treating years 1 and 2 as baseline years (the 'seven-year baseline' model). The second model restricted the baseline period to two years, EC 2002 and EC 2003, and used a 'difference-in-differences' approach, a standard estimating technique in program evaluation. The two models are described and their estimates presented in Appendix 4.

<sup>32</sup> In cases like the present, where a large number of impact estimates are calculated, the t-statistic required for statistical significance can be higher.

<sup>33</sup> The evaluation team considered the calculation of minimum detectable effects but decided against doing so for the reasons explained in Appendix 4.

<sup>34</sup> Statistical significance was tested using the Benjamini-Hochberg adjustment for multiple comparisons. For a succinct explanation of this procedure see Peter Z. Schochet, *Technical Methods Report: Guidelines for Multiple Testing in Impact Evaluations* (NCEE 2008-4018), Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education, 2008.

Neither alternative model yielded statistically significant impacts on the number of sitters, for either boys or girls, in any year of the pilot. Moreover, both of these alternative models showed substantially smaller upper bounds on the impact than those shown in Table 2.1. The seven-year baseline model yielded upper bounds of 18 per cent for boys and 19 per cent for girls in the emerging regions, 10 per cent for boys and 6 per cent for girls in the non-emerging regions, and 12 per cent for boys and 7 per cent for girls nationally. The difference-in-differences model, which could not be estimated in the emerging regions, showed upper bounds of 3 per cent for boys and -6 per cent for girls in the non-emerging regions. Again, it should be noted that there is only a 5 per cent chance that the true impact was as large as the upper bound of the 95 per cent confidence interval. Appendix 4 provides the estimates of the upper bound of the 80 per cent confidence interval; true impacts have a 20 per cent chance of being as large as this bound.

## 2.2 Impacts on the Numbers of Passers Attributable to the RBA Pilot

The number of students passing the EGSECE is simply the number sitting for the examination multiplied by the pass rate (the proportion of students with GPAs above 2.0). Thus, any difference in deviations from trend between the number sitting and the number passing is attributable to changes in the pass rate.

As noted in the evaluation team's earlier reports, the number of students passing the EGSECE is not a good measure of educational progress for several reasons. First, the NEAEA normalises scores on the examination every year. This means that changes in the mean (raw) scores on the examination have little or no effect on the pass rate, as explained in Chapter 1. In fact, *under this system, the pass rate should be relatively unchanged from year to year*. As the pilot's verification team observed, 'Current procedures ... that involve normalisation of raw scores at the subject level are explained in detail in the verification baseline report, where it is demonstrated that the pass rate is expected to be in the range from 67 per cent to 70 per cent', regardless of educational performance<sup>35</sup>.

Second, even if normalisation did not eliminate the changes in the pass rate from year to year – and there are in fact substantial changes in that rate – it would be impossible to distinguish changes that reflect improvements in educational performance from those that simply reflect changes in the difficulty of the examination. The EGSECE uses new questions every year, and little attempt is made to calibrate the difficulty of the questions to make them comparable over time<sup>36</sup>. Instead, normalisation is used to even fluctuations in the difficulty of the examination.

More important for present purposes, the scoring procedures for the EGSECE changed fundamentally in EC 2002. Prior to that year students were required to receive a passing grade in both English and mathematics. Since EC 2002, they have been required to receive only a GPA of 2.0 or greater averaged across their seven best subjects including English and mathematics to pass. In part as a result, the pass rate was low and fluctuated erratically prior to EC 2002. In 2002, when these changes were made, the pass rate jumped 20 points and since then has hovered in the range of 60-70 per cent. This change in procedures – and break in the trend line – means that the CITS model cannot be used to estimate impacts on the number of students passing the examination.

<sup>35</sup> Coffey International Development, *Phase 4 Verification Report, DFID Ethiopia Independent Verification of Key Government of Ethiopia (GOE) Educational Data for a Pilot of Results-Based Aid*, 2015. An independent analysis of the pass rate (available from the authors on request) also suggests a virtually constant pass rate over a fairly wide range of mean raw scores.

<sup>36</sup> It is important to note that this is not a problem for the purposes for which MoE uses the EGSECE – determining students who pass grade 10 and students who are allowed admission to upper secondary school. These purposes do not require comparisons across time; in fact, for these purposes, a relatively stable pass rate is desirable.

Although the number of students passing the examination is not a useful measure of educational progress, the analysis nevertheless estimated RBA impacts on that measure because it was used as the pilot's payment criterion. The change in grading procedures in EC 2002 precluded the use of the CITS model, so for these estimates a 'difference-in-differences' model that relied on data only for the years EC 2002-2006 was used<sup>37</sup>. Difference-in-difference is the method used in most non-experimental evaluations in which longitudinal baseline data are unavailable – as it is in most evaluations. This model relies on non-government schools as a comparison group, though, so it cannot be estimated for the emerging regions, which have too few non-government secondary schools. Table 2.2 shows the estimated impacts on the numbers of boys and girls passing the EGSECE in the non-emerging regions and nationally. As in the previous table, the estimates are expressed as percentage differences from the value that would have existed in the absence of the RBA pilot.

Table 2.2: Impacts, Years 1-3, Numbers of Boys and Girls Passing the EGSECE

	Boys			Girls		
	Year 1	Year 2	Year 3	Year 1	Year 2	Year 3
<b>Non-emerging regions</b>	-10.4% (-1.9)	-6.3% (-1.17)	-23.5% (-4.32)	-18.7% (-2.95)	-21.4% (-3.36)	-44.4% (-6.92)
	[-1.4%]	[2.5%]	[-14.6%]	[-8.3%]	[-11.0%]	[-33.9%]
<b>National</b>	-8.2% (-1.49)	-5.1% (-.94)	-17.7% (-3.22)	-18.2% (-2.68)	-20.6% (-3.03)	-39.5% (-5.78)
	[0.8%]	[3.8%]	[-8.7%]	[-7.1%]	[-9.5%]	[-28.3%]

Note: The t-statistics for each estimate are in parentheses. The upper bound of the 95 per cent confidence interval (one-tailed test) is shown in square brackets below the t-statistic.

Source: Regression analysis of data from the MoE and the NEAEA, EC 2002-2006.

As can be seen, the estimates in Table 2.2 are uniformly negative, and the upper bounds of the 95 per cent confidence intervals are all either negative or negligibly small. Although some of these estimates are large and significantly different from zero, negative estimates are not taken as evidence of the pilot's impact for several reasons. First, such impacts would be illogical – there is no reason to expect payments for increased numbers of passers to reduce the number of students passing the examination. Second, there is little reason to expect impacts on numbers passing the EGSECE in years 1 and 2 because the RBA scheme had not been communicated to the regions in time to affect school performance. Third, as noted above, there is good reason to believe that fluctuations in the pass rate simply reflect the nature of the grading system, in particular the fact that NEAEA uses new questions every year and normalises the examination scores.

In conclusion, use of the CITS model *did not detect evidence that the RBA pilot improved educational performance for either boys or girls, in either the emerging regions or the non-emerging regions, in any year of the pilot.*

The reason for lack of any detectable effect (beyond the measurement and estimation challenges) is not difficult to discern. As discussed elsewhere in this report and as just mentioned, the RBA scheme had not been communicated effectively to the regions in time to affect students' performance appreciably during the pilot's three years. When the evaluation team visited regions in years 1 and 2 of the pilot, virtually none of their education officials were aware of it. When the verification team surveyed schools following the EC

<sup>37</sup> See Appendix 4 for a detailed discussion of this model.

2006 EGSECE, at the end of year 3, only about half the head teachers they interviewed knew about the pilot<sup>38</sup>.

In addition, given the nature of the grading system, there is no reason to expect RBA to have any effect on the number of students passing the examination other than through its effect on the number sitting for it. For this reason, the evaluation team recommends that this indicator not be used in the future as a criterion for RBA payments.

There is, however, some reason to believe that RBA could affect the number of students sitting for the examination if the pilot is continued. As detailed in subsequent chapters, RBA funds have begun to flow to regions and schools and, as a result, many regional and local staff are now aware of the RBA incentive. Improvements to the educational system are being planned and some are being implemented (see Chapter 5). These improvements may well affect the number of students sitting for the EGSECE in future years.

---

<sup>38</sup> Coffey International Development, *Phase 4 Verification Report, DFID Ethiopia Independent Verification of Key Government of Ethiopia (GOE) Educational Data for a Pilot of Results-Based Aid*, 2015.

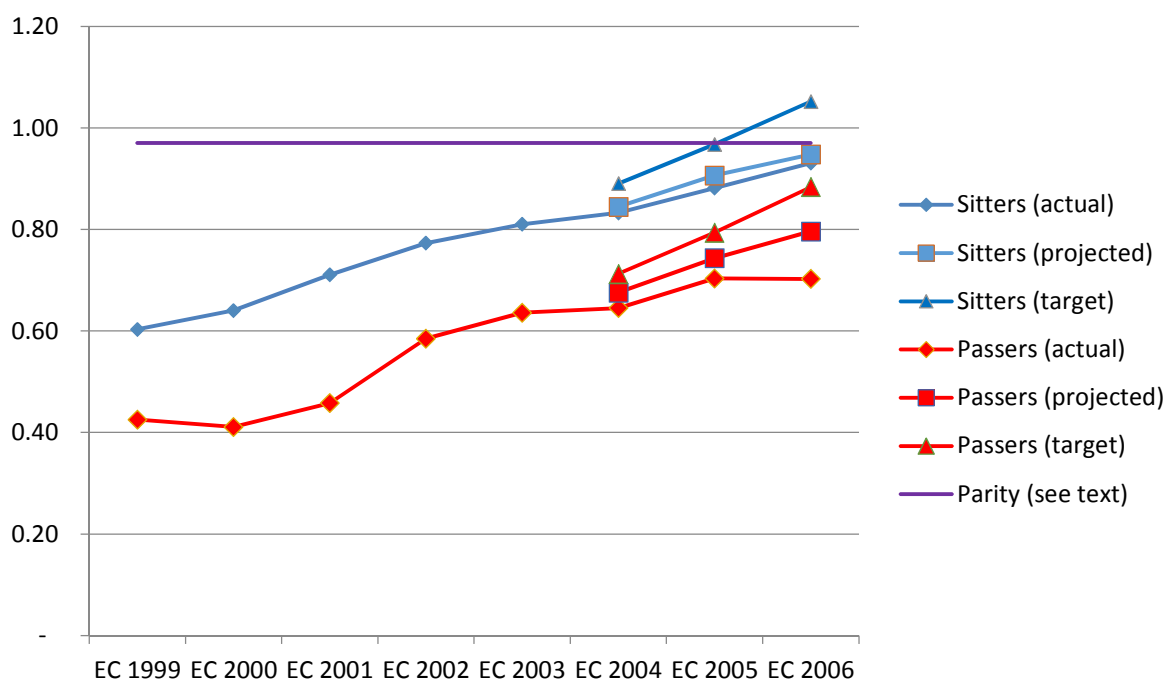
## 3 The RBA Pilot's Contribution to Gender and Regional Equity

### 3.1 Gender Equity

The Gender Parity Index (GPI) measures parity between girls' and boys' participation in education. The index represents the ratio of female-to-male values of a given indicator. Assuming an equal number of boys and girls, a GPI of 1.00 would indicate parity between females and males. A value less than 1.00 indicates a disparity in favour of boys. To illustrate, a value of 0.50 means that girls enjoy half the access that boys do. A value greater than 1.00 indicates a disparity in favour of girls. In Ethiopia a GPI of 0.97 indicates gender parity in lower secondary school nationally. At that point the enrolment, sitting, and passing ratio would match the population ratio of girls to boys, and girls would have the same chance as boys of enrolling, sitting, or passing the EGSECE. The corresponding parity value in emerging regions is 0.75, while in other regions it is 1.00.

Figure 3.1 shows that the GPI for EGSECE sitters and passers increased markedly over the three years of the pilot. Despite these increases, they did not meet the expectations associated with maximum fund disbursement that DFID had included in its *Business Case*, which anticipated extra progress for girls compared to boys. Neither did the GPIs in the three pilot years increase by as much as the evaluation team had projected in its baseline report. These projections are based on the results of previous examinations and data on enrolment by grade in the absence of an RBA pilot. Indeed, the ratio of female-to-male passers did not increase at all in EC 2006 because the increase in girl sitters was offset by the greater decline in pass rates they suffered. In sum and most important, the lack of gap-closing beyond the expected projected trends indicates there is no evidence that the RBA pilot has accelerated progress toward gender equity. On the contrary, the increases that have occurred are largely a continuation of pre-existing trends rather than a change attributable to the RBA pilot.

Figure 3.1: Gender Parity Indices for EGSECE Sitters and Passers





Fifty-two per cent of female candidates passed the examination in EC 2006, but this was only *three quarters* of the male pass rate of 69 per cent, roughly in line with the ratio since EC 2002. Furthermore, looking only at the pass rate understates the gender gap. Girls are only *half* as likely as boys to achieve a grade point average of 3.0 or more (rather than a pass mark of 2.0) out of a maximum of 4.0, which is important. Only students with the best marks are selected to enter upper secondary school.

The reasons for the large gap between girls' and boys' success in the EGSECE are not fully known. The gender gap in the pass rate has been similar in grade 12 since EC 2002 at 16 to 17 percentage points, but much narrower at grade 8 at the end of primary schooling, averaging only two percentage points since EC 2004. The National Learning Assessment in 2010 confirms large gender gaps at grades 10 and 12 across subjects and regions<sup>39</sup>. As the recent evaluation of GEQIP concluded, 'gender disparities which might impact on learning opportunities for either boys or girls could not be observed'<sup>40</sup>. Two recent research papers point to puberty as a key factor in Ethiopia and link menstrual cycles with girls' non-attendance for – or poorer concentration during – lessons and examinations<sup>41</sup>. Some schools are constructing appropriate, separate toilet blocks for girls. Sensitising boys and girls to menstruation and its effects may also be important. Sexual harassment by staff, lengthy journeys to school, domestic duties, early marriage, and a large gender disparity among teachers may also make it harder for girls to achieve the same learning outcomes as boys, especially in rural areas.

Results from the impact modelling, described in the previous chapter, are consistent with these findings. There were no statistically significant, positive deviations from trend for either boys or girls during the pilot's three years, suggesting an uninterrupted continuation of the trends observed in the baseline period for outcome measures for both males and females. It must be borne in mind that the increases in girl sitters from EC 2004 to EC 2006 were driven by equally large enrolment increases, and that the impact model controls for changes in enrolment. *There is no evidence from the impact model discussed in the previous chapter that the RBA pilot caused or contributed to a narrowing of the gender gap in sitting or passing the EGSECE in EC 2004, EC 2005 or EC 2006.*

The lack of evidence for an RBA effect in favour of girls sitting or passing the EGSECE is similarly consistent with the fact that RBA messages, incentives, and resources did not flow in time to influence year 3 results. The evaluation team did not identify any major actions taken in response to incentives prior to fund disbursement. Furthermore, the evaluation team's field-based research did not find much of a gender dimension in the use of RBA funds that were disbursed. The construction of a cafeteria for girls in one school in Somali, the plans to construct a separate toilet block for girls in one school in Oromiya, and the celebration of female success in the EGSECE in Gambella represent minor exceptions.

The evaluation team's school-level fieldwork identified various interventions unrelated to the pilot, such as community outreach, facility construction, and extra lessons for girls that are designed to address factors that make girls less likely than boys to stay in school and sit or pass the EGSECE. The fieldwork suggests that such initiatives can be successful, but also that they can be overwhelmed by social, economic, and even geographical factors, including negative attitudes to schooling, the need for students to take paid employment, and the distance some students must travel to school, especially in the emerging regions.

<sup>39</sup> NEAEA, *Ethiopia First National Learning Assessment of Grades 10 and 12 Students*, December 2010.

<sup>40</sup> HIFAB International, *Comprehensive Evaluation of the General Education Quality Improvement Programme (GEQIP) – Draft Exit Survey Report*, May 2013.

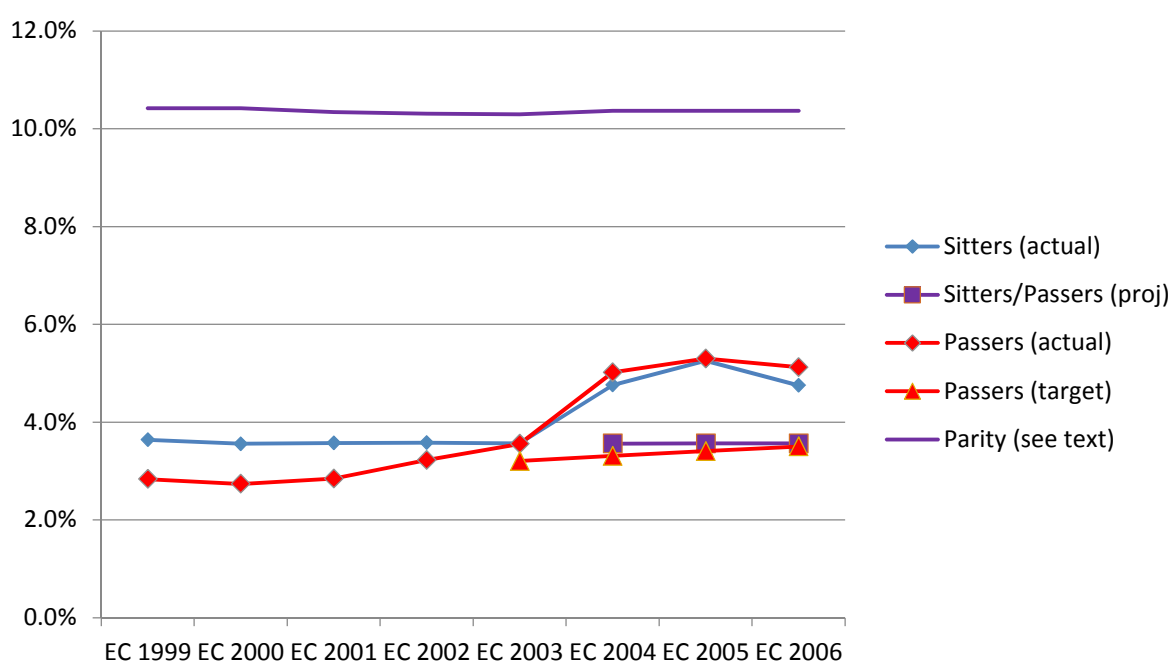
<sup>41</sup> Teketo Tegegne and Miteke Sisay, 'Menstrual hygiene management and school absenteeism among female adolescent students in Northeast Ethiopia', October, 2014. Available at [www.biomedcentral.com/1471-2458/14/1118](http://www.biomedcentral.com/1471-2458/14/1118), and Teklemariam Gultie, Desta Hailu and Yinager Workineh, 'Age of Menarche and Knowledge about Menstrual Hygiene Management among Adolescent School Girls in Amhara Province, Ethiopia: Implication to Health Care Workers & School Teachers', September 2014. Available at <http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0108644>.



### 3.2 Regional Equity

Slightly more than 10 per cent of all 15 and 16 year olds in Ethiopia are estimated to live in the four emerging regions. Regional parity would be achieved when a similar percentage of Ethiopia's sitters and passers are from these four regions. As shown in Figure 3.2, however, such parity has not yet been achieved. At about 5 per cent, the share of sitters and passers in the emerging regions was higher in all three pilot years than the share in previous years (i.e., slightly less than 4 per cent). This increase only marginally addressed the disparity between the emerging and non-emerging regions.

Figure 3.2: Regional Equity: The Percentage of Sitters and Passers in Emerging Regions – Actual, Projected and Targets



Further analysis presented in Appendix 5 suggests the levels and trends differ considerably among the emerging regions. In the five years up to EC 2003 children in both Afar and Somali were roughly six times less likely to sit the EGSECE than children in non-emerging regions, which was about one in three. The proportion remained the same in Afar and improved in Somali to one in three in EC 2005 and 2006. Since EC 1999, children in Benishangul-Gumuz had at least the same chance as those in non-emerging regions, while children in Gambella have had a greater chance, rising to double in EC 2005 and 2006 at around two in three. The extra resources offered under the pilot for additional sitters in emerging regions have a clear rationale with respect to Afar and Somali, but not for Gambella or Benishangul-Gumuz.

The emerging regions' share of sitters and passers declined in EC 2006, but candidates in emerging regions were more likely to be successful (66 per cent pass rate in EC 2006) than their counterparts elsewhere (61 per cent), which is unusual historically. Further analysis of individual regions indicates a mainly inverse relationship between sitting for the examination and attainment in the emerging regions in recent years (see Appendix 5). Gambella had the lowest pass rate of any region (41 per cent in EC 2006), Benishangul-Gumuz was well below average, Afar was slightly above average, and Somali had the highest pass rates of any region (81 per cent in EC 2006).

The share of sitters and passers exceeded the expected results in DFID's *Business Case*, though it should be noted that these results did not allow for emerging regions to close the gap with the non-emerging regions. The results also exceeded the evaluation team's conservative projections in the baseline report<sup>42</sup>. The analysis indicates that a major construction programme for lower secondary schools in emerging regions, notably in Somali, led to increases in enrolment and sitters above what would have been expected based on trends from previous years. The increase in sitters in EC 2004 coincided with large declines in sitters in Oromiya and Amhara, whereas the smaller increase in emerging regions in EC 2006 was outweighed by the increase in other regions.

The impact model, discussed in Chapter 2, found no evidence of statistically significant deviations from baseline trends in the numbers of students sitting or passing the examination in either the emerging or non-emerging regions. It can be concluded, therefore, that while emerging regions reduced the gaps during the pilot, the gains were unlikely due to the pilot.

As with gender equity, the regional gap-closing in the first two years of the pilot predates any action detected in response to pilot incentives or the use of RBA funds. Long-standing problems (including student disengagement, poor facilities, and inadequate resources) are typically worse in the emerging regions, and these problems are not easily, quickly, or inexpensively mitigated.

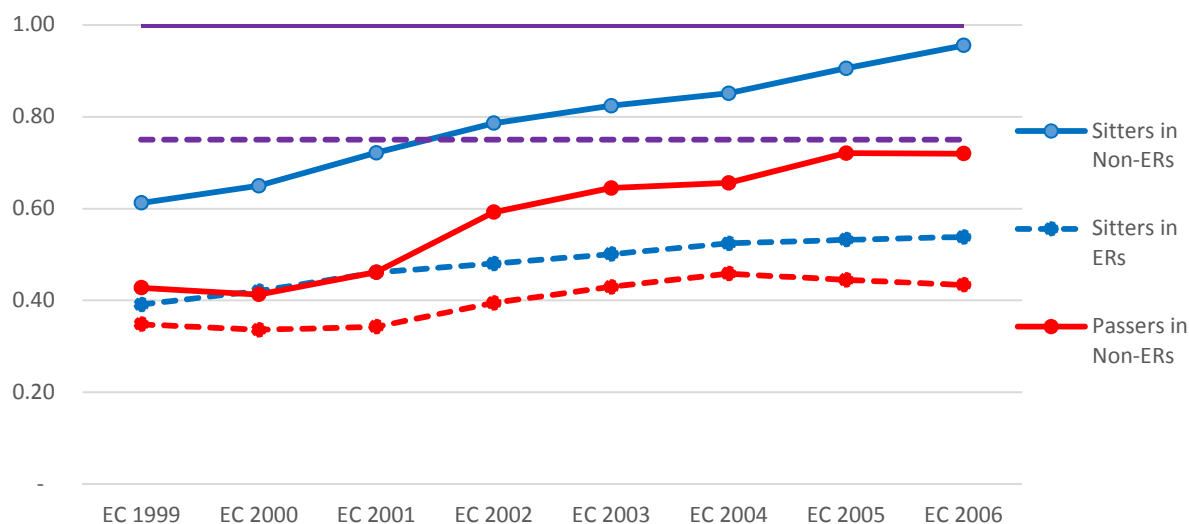
### 3.3 Gender and Regional Equity Combined

Figure 3.3 charts the GPIs for emerging and non-emerging regions separately and suggests some interaction between gender and regional equity over the period of the pilot and in the preceding years. The trends for non-emerging regions largely repeat the national gender trends because they represent the majority of the national figures. Thus gender parity in non-emerging regions was nearly achieved for *sitters* by EC 2006 (the GPI rising to 0.96 by EC 2006 from 0.82 in EC 2003 compared to parity at 1.00 for non-emerging regions), but not for passers. The pass rate for girls continued to be below that for boys for all years from EC 1999 onwards (GPI reached only 0.72 in EC 2006). By contrast, the GPI for sitters in emerging regions increased more slowly during the pilot years (rising to 0.54 from 0.50 compared to a parity value of 0.75)<sup>43</sup>. Again, girls remain less likely to pass the EGSECE than boys in emerging regions than elsewhere, and the GPI for passers for EC 2006 was no higher than the GPI three years earlier (i.e., 0.43 in EC 2003).

<sup>42</sup> Data from the MoE's education management information system suggested that an above-average rise in EGSECE sitters might follow, but the erratic series and some known problems with the data encouraged a cautious approach to the assumptions associated with the projections.

<sup>43</sup> Note that the parity target is only 0.75, reflecting the considerably greater ratio of 16-year old boys to girls in emerging regions. It is possible that some girls are moving from emerging regions to other regions to access lower secondary education. If this is occurring, which is unlikely to be on any great scale, then girls' access in emerging regions would be understated and their access in the other regions would be overstated.

Figure 3.3: Gender Parity Indices for EGSECE Sitters and Passers in Emerging Regions (ERs) and Other Regions (Non-ERs)



The RBA pilot offered additional incentives for emerging regions and for girls, especially those in emerging regions, to sit for and to pass the EGSECE. Despite these additional incentives, the widening gender gap in emerging regions compared to the rest of Ethiopia indicates that the pilot has not successfully tackled long-standing patterns of gender inequity. This unfortunate result may be due to the lack of action in response to the RBA incentives in time to affect the EC 2006 data. Another reason for the lack of progress by girls in emerging regions may be the change made by MoE in the formula to allocate half of the resources based on regions' total numbers of sitters and passers, rather than wholly according to results achieved. Based on the MoE's original formula for disbursing the RBA funds, emerging regions 'earned' £1.8 million in the first three years but received just under £1.4 million as a result of the MoE's revisions in the allocation formula<sup>44</sup>. Somali actually received about £418,000 less than its sitters and passers had earned.

### 3.4 Other Disparities

Girls and young people in emerging regions are the two groups identified as suffering inequity and their prioritisation is embedded in the pilot's design. Other disadvantaged groups can be identified, such as children from poor households, children with disabilities, and children in remote pastoralist communities. No data were available for these groups, so it is not possible to draw any conclusions about changes in equity for them.

### 3.5 Conclusions

Drawing the strands of the analysis together and expressing them through the lens of the theory of change, the evidence suggests that existing trends and factors *outside* the RBA pilot have contributed to the gender and regional gap-closing in sitting and passing the EGSECE. The gender gap appears to have closed

<sup>44</sup> As noted in Chapter 1, the MoE decided to allocate half of each year's reward payments received from DFID to regions according to the number of sitters and passers (i.e., their size in lower secondary terms). The four emerging regions therefore received only 75 per cent of the reward payments they would have received for EC 2004, 2005, and 2006 had the allocations been based solely on their increases in the number of sitters and passers in those years.

more slowly in emerging regions than elsewhere, and female EGSECE sitters continue to be considerably less likely than males to score high marks.

The absence of evidence indicating that the pilot has had an impact on the numbers of sitters and passers makes it difficult to comment on whether the premiums in the national reward formula – and hence regional allocation – for girls and emerging regions for sitters and passers have been set at the ‘right’ or optimal amount (and as also discussed in Chapter 7). The data on the current gender and regional gaps for sitters and passers may, however, be combined with the information on how regions have spent the initial funds to suggest alternative reward premiums (see Table 3.1). If gaps have closed, there is no evidence that the reward payments have been effective or that the pilot has contributed to closing any gaps. If a gap remains and the premium has not led to actions that might close that gap in the future, a larger premium might have been more effective. If the pilot is continued beyond year 4, the evaluation team recommends that the premiums be reviewed based on these comments.

Table 3.1: Proposed Alternative Premiums in the Allocation Formula

Group	Premium (compared to boys in non-emerging regions @ £50)	Comments on premium
Sitters: boys in emerging regions	+ 50 per cent (£75)	Possibly too low for Somali and Afar, where access remains low, but unnecessary for Gambella and Benishangul-Gumuz, where access is greater than in non-emerging regions
Sitters: girls in non-emerging regions	+ 70 per cent (£85)	May be unnecessary as girls almost equally likely to sit EGSECE as boys in non-emerging regions (GPI = 0.96, see Figure 3.3)
Sitters: girls in emerging regions	+ 100 per cent (£100)	Too low as large gender gap in access in emerging regions remains (Figure 3.3) and limited evidence of pro-girl use of RBA funds to address this gap
Passers: boys in emerging regions	+ 50 per cent (£75)	Perhaps unnecessary as pass rates are now higher in emerging regions than elsewhere (though they may fall in emerging regions if access increases in the future)
Passers: girls in non-emerging regions	+ 70 per cent (£85)	Too low as large gender gap in pass rate remains and limited evidence of pro-girl use of RBA funds to address this gap
Passers: girls in emerging regions	+ 100 per cent (£100)	

## 4 The Systems Effects of the RBA Pilot

According to DFID's *Business Case*, the key assumption of effectiveness of RBA is that it incentivises change in country systems. In addition, the pilot, according to DFID, would promote the government's accountability to its citizens and establish the basis for improvements in public financial management in the education sector. This chapter addresses these issues as well as the validity of DFID's assumptions about the systemic effects of the RBA pilot.

In principle RBA can strengthen or weaken country-based systems and their usage. Positively, RBA could encourage increased attention to data on outcomes, on analysing them, and supporting sources to design actions likely to achieve rewarded results. In turn this attention could help lead to improved data accuracy and timeliness. Negatively, the temptations for seeking shortcuts to financial rewards through misrepresentation could undermine the results system, hence the need to verify the results reported. In the present instance the evaluation team sought to determine whether the channelling and use of the RBA payments has supported or undermined national resource allocation processes and outcomes, education management information systems (EMIS), and financial management systems. This was done through the analysis of trends in data timeliness, data quality, and resource allocation before and during the pilot and through interviews with education officials in Ethiopia.

Given the findings noted in previous chapters, one would not expect to see major or perhaps even discernible effects on the underlying systems by the end of EC 2006. *In fact, the evaluation team is not aware of any action taken within the system in response to incentives to obtain greater rewards in advance of the funds being disbursed.* The RBA rewards began to flow only at the latter stages of the pilot, with the consequence that there was limited opportunity to influence systems during the pilot's first three years. Moreover, there has not been a strong focus on fostering enhanced accountability for results or on scrutiny by stakeholders.

### 4.1 Resource Allocation Outcomes

DFID hypothesised in its *Business Case* that RBA would result in no change in the share of resources devoted to secondary education. This is a difficult hypothesis to test; such an outcome would be consistent with absolutely nothing happening. Unfortunately, the evaluation team was not able to obtain the data necessary to test the hypothesis and is, therefore, unable to make any definitive statement about whether a meaningful reallocation of resources occurred. Even if it had been possible to identify allocation patterns, the large changes in allocations to secondary education during this period and the relatively low share of RBA payments as a proportion of overall spending on regional recurrent expenditure on education – less than 1.2 per cent in year 2 – make it extremely likely that allocations would not have changed.

### 4.2 EGSECE Administration and the MoE's EMIS

The verification team's initial assessment of the robustness of the EGSECE's administrative systems and its recurring assessments of the data in a sample of schools all provide reassurance that the system was sound at the outset and continues to be<sup>45</sup>. The evaluation team recommended that DFID commission the verification team to repeat its surveys of examination invigilators and teachers that were conducted for that team's baseline report. As this suggestion was not agreed it is not possible to offer a view on whether the administration of the EGSECE has seen any improvements, remained the same, or suffered any deterioration. The NEAEA has continued to administer the EGSECE in the context of the pressures of a

<sup>45</sup> Coffey International Development, *Inception Report, Baseline Report, and Phase 2, 3 and 4 Verification Reports: Independent Verification of Key Government of Ethiopia (GoE) Educational Data for a Pilot of Results-Based Aid (RBA)*, January 2012, July 2012, January 2013, December 2013, and January 2015.

national high-stakes examination, which are likely to be greater than any the RBA pilot could have generated, so the pilot is unlikely to have caused or contributed to any meaningful changes that may have occurred.

In a review of the RBA pilot, the Center for Global Development (CGD) noted that the EGSECE made it difficult to assess any changes in student learning and argued that the 'pilot provides an ideal setting for the government of Ethiopia to set in place a process for establishing a test that meets a global standard in terms of equivalence'<sup>46</sup>. Despite this perceived opportunity, the MoE and NEAEA have not changed the EGSECE. The switch from norm-referenced to criterion-referenced examination that would have been required is a major undertaking needing substantial investment over a lengthy period. Indeed the pilot made it *harder* to reform the examination reform, because it would not be possible to calculate the reward payment for passers in the first year of any new system (and CGD made its recommendation after the pilot baseline had been conducted and validated).

There is no reason for the pilot to undermine the quality of the school-reported EMIS data. Altering these data would not affect the results that determine the size of the payments. The timeliness of the EMIS, as judged by the release dates of the *Education Statistics Annual Abstract*, which includes the annual results, was much slower in EC 2006 than in the two previous years. The EC 2006 abstract was issued on June 30, 2015, nearly eight months later than the corresponding abstract for EC 2005 (November 2013) and nearly a year longer than the abstract for EC 2004 (September 2012). Why has this delay occurred? The departure of key staff appears to be the cause rather than the pilot. It is difficult to assess changes in the quality of the data before EC 2006, but no pilot-related improvements were detected. There were problems obtaining data from Somali in EC 2002. They have not recurred, but the problems were likely due to the changing security situation and the MoE's efforts independent of the RBA pilot. By the same token, the reduced number of returns from non-government schools in Addis Ababa in EC 2005 does not appear to be related to the pilot.

### 4.3 Financial Management System and Resource Allocation Processes

In principle, RBA can be used in any sector of the national system, as the government sees fit. If RBA rewards are transferred to the MoFED without being earmarked they could subsequently be distributed alongside or comingled with any other financial resources. In the RBA pilot, while RBA rewards were channelled through MoFED they were earmarked to the MoE and resources were subsequently released to the regions according to the MoE's revised guidelines.

In terms of the use of RBA reward payments MoE could, if it had so wished, have simply channelled resources through national processes. Instead, it chose to establish a dedicated allocation, funds transfer, planning and reporting system – as set out in its guidelines – to disburse the RBA payments to the regions. It is not clear, however, whether this was entirely an MoE decision or whether this approach was taken because MoE officials believed that this was what DFID expected. In this instance, DFID's *Business Case* is not clear. On the one hand, the *Business Case* declares that the MoE *in collaboration with MoFED* would be 'responsible for managing the allocation of the additional resources to the eleven regions'. On the other hand, the *Business Case* also notes an agreement to channel the additional payments through MoE *rather than through MoFED* because this was judged to be more likely to generate incentives for action in the education sector. The first option was chosen.

<sup>46</sup> Nancy Birdsall and Rita Perakis, 'Cash on Delivery Aid: Implementation of a Pilot in Ethiopia', April 27, 2012. Available at [http://www.cgdev.org/sites/default/files/archive/doc/Initiatives/Ethiopia\\_RBA\\_pilot\\_report.pdf](http://www.cgdev.org/sites/default/files/archive/doc/Initiatives/Ethiopia_RBA_pilot_report.pdf)

In sum, while the delivery and use of the RBA payments did not conform to existing national processes and undoubtedly increased transaction costs, the government supported or at least tolerated the deviation from such processes in the expectation of some benefits from the arrangement. Indeed, there is a strong case for not integrating innovative approaches within existing and familiar systems if there are concerns about the sustainability of the approaches, and DFID's own guidance on aid instruments makes this case<sup>47</sup>.

#### 4.4 Accountability

Structures are in place to ensure accountability of the RBA process but the limited activity within the regions (see Appendix 6) by the end of year 3 meant there was little evidence of the structures' use. The MoE's guidelines had required REBs to submit regional action plans describing how the rewards payments would be used, but few REBs did so.

Appointed RBA leads within the REBs typically developed the plans. All of the completed plans corresponded with their regional Growth and Transformation Plans and Regional Improvement Plans and so meshed with previously sanctioned processes. Among the plans that were submitted, however, the MoE provided limited scrutiny or review.

In the regions in which schools were given responsibility for spending RBA funds, schools were also required to submit plans to their REBs. These plans were typically agreed by a council comprising the head teacher, teacher and student representatives, and one or more members of a PTSA. In some cases, representatives of woreda education offices (WEO) advised on planning<sup>48</sup>. The school plans reflected their own school improvement plans and so also complemented previously sanctioned processes. There appears to have been strong local accountability in instances in which schools were given responsibility for spending RBA funds.

The MoE's guidelines required REBs to report to the MoE how the regions' funds had been spent and how the expenditures corresponded to the submitted plans. Despite this requirement, most REBs struggled to spend their reward payments in the time frame that MoE had established and no reports had been submitted to MoFED by the end of year 3. Likewise, schools given responsibility for spending RBA funds were also required to submit a full accounting of how they had spent their rewards. Again, however, there was no evidence, at least observed by the evaluation team, that schools had submitted their reports by the end of year 3. What the regions and their schools did with their reward payments is discussed in Chapter 5.

In sum, the structures in place to ensure accountability of the RBA process, including financial reporting, corresponded with extant structures. That is, no new structures were put in place. Particularly at the local level, there were some opportunities to demonstrate accountability through the use of these structures. Nonetheless, the relative lack of RBA activity at the local and regional levels during the pilot meant that most of these structures have not been tested or their effectiveness evaluated.

#### 4.5 Corruption

Identifying corruption in the use of RBA money would have been possible only with forensic accounting skills outside the scope of this evaluation. As was argued in the evaluation team's prior reports, if corruption had occurred the problem was less the corruption itself but more an issue of the reward levels

<sup>47</sup> DFID, 'Guidance on Aid Instruments: A DFID Practice Paper', n.d. Available at <http://www.mtnforum.org/sites/default/files/publication/files/1873.pdf>

<sup>48</sup> A woreda is a local administrative division within Ethiopia and is equivalent to a district.

that allowed the MoE to misuse funds yet still achieve the necessary results. Furthermore, as Charles Kenny and William Savedoff have argued, 'the mechanisms used to measure outcomes (or outputs) in results-based modalities necessarily reduce the scope for corruption'<sup>49</sup>.

---

<sup>49</sup> Charles Kenny and William Savedoff, 'Can Results-based Payments Reduce Corruption?' CGD, Working Paper 345, 2013. Available at <http://www.cgdev.org/publication/can-results-based-payments-reduce-corruption-working-paper-345>. They further argue that this is because 'when outcomes are under-priced, individuals who wish to defraud the program have no incentive to participate. Even when outcomes or outputs are overpriced, those who wish to divert funds from a program can only do so by generating results'.



## 5 Understanding Why the Approach Did or Did Not Work

As the analysis in Chapter 2 concluded, there is no statistically significant evidence that the RBA pilot contributed to increases in the numbers of EGSECE sitters and passers. There is some evidence, however, of RBA-related initiatives among regions and schools that may lead to increases in the numbers of sitters and passers in the future.

This chapter considers the pilot's potential for success and then examines the factors that mitigated that success during the pilot. The chapter draws on the qualitative research conducted for the evaluation. This research involved interviews with the MoE and REBs as well as with head teachers, teachers, students, district-level officials and members of PTsAs in selected schools across the 11 regions. In keeping with the widely accepted practices of qualitative research, the process-tracing data are illustrative and context-based. The interviews were semi-structured. That is, they focused on a set of key issues relating to RBA (to ensure the consistency of the questioning) while also giving research participants the opportunity to address individual and particular issues (to ensure proper attention was given to context-specific details). This approach allowed continual validation of the data: clarifications were sought during the interview if necessary, and the responses were summarised at the end to ensure the accuracy and fairness of the record.

The national- and regional-level research was comprehensive in that key stakeholders in the MoE and all REBs were interviewed. The fact that all REBs were interviewed is important because all the reward payments that were distributed to schools first went to the REBs, and each of them devised their own approaches to the intraregional distribution and use of the reward payments. As a consequence, the qualitative research provides a thorough review of regional decisions about and reactions to the RBA pilot, and much of the discussion that follows relies on the comprehensive qualitative data from the national and regional level.

In contrast, school-level data were more limited. Thirty schools across the country were visited as part of the baseline study (year 0 and year 1), so the ability to generalise the findings to all schools in Ethiopia is thus limited. The evaluation team intended to visit 33 schools (three in each region) in year 3. This effort was halted after nine school visits because the team realised how little RBA activity was occurring within schools. In the near-universal absence of awareness of the pilot, it made little sense to continue with the visits. The value for money in completing more school visits was negligible. Further information on the methods of data collection and the conduct of the research can be found in Appendix 6, which also includes summaries of regional and school-level activities.

### 5.1 What Happened

At the national level, the MoE took initial responsibility for RBA. The MoE then devolved responsibility to the regions to address the diversity of their needs. When it was realised that the total year 1 reward would be small, the MoE negotiated with DFID to: (a) change the formula for rewarding regions so that regions making no progress would remain engaged with the pilot; and, (b) combine the year 1 and year 2 payments to the regions to give greater impetus to their engagement with the RBA pilot.

The announcement of the combined regional rewards was made at the annual education conference in October 2013<sup>50</sup>. The MoE then sent its RBA guidelines to the REBs and asked them to develop regional

---

<sup>50</sup> This timing meant that two of the three examinations during the pilot had already been completed (in May/June 2012 and May/June 2013) by October 2013.

plans for spending their rewards. Delays occurred in disbursing the rewards to the regions. The MoE and MoFED first had to reach agreement on how payments would be made to the regions. This agreement was reached in April 2014 (EC 2006). Due to the delay some REBs did not receive their funding until or after the end of the school year and thus after the third and last administration of the EGSECE during the pilot. By the end of year 3, however, some REBs had spent or were spending their rewards, and there was some evidence of RBA-funded changes at the school level.

Two main approaches to RBA emerged in the regions during year 3 (see Table 5.1):

- the school approach, in which REBs disbursed funds to schools for school-level activities; and,
- the regional approach, in which REBs retained funds for regional activities.

Three variations of the school approach were also apparent:

- support for all schools (although amounts or resources allocated to schools varied);
- support for high-achieving schools, in which most funds were allocated to schools with the best results; and,
- support for underresourced and/or underachieving schools.

Some REBs adopted a mixed approach, for example, by retaining some funds for regional activities and disbursing other funds to high-achieving or underresourced schools.

Table 5.1: Regional Approaches to RBA

Approach	Support for		
	All schools	High-achieving schools	Underresourced schools
Regional	Addis Ababa*		Amhara*
	Afar		Benishangul-Gumuz*
	Benishangul-Gumuz*		Gambella*
	Gambella*		Somali
	Harari		
School	Dire Dawa	Addis Ababa*	Tigray
	SNNPR	Amhara*	
		Oromiya	

\* denotes a mixed approach to RBA.

Regional- and school-level RBA action plans reflected their existing development plans, including regional Growth and Transformation Plans, but the limited reward payments required REBs and schools to prioritise their spending. The delays in making payments to REBs and schools (for those regions employing the school approach) inhibited RBA activity across the country. There were indications, however, that the RBA process did facilitate changes in thinking at the regional level, as discussed below. There was also some evidence of strategic thinking among schools that received funds from their REBs, but this interpretation is necessarily limited due to the small number of schools the evaluation team visited in year 3.

An appointed RBA lead typically made initial decisions at the regional level and then the REB's head and/or a representative REB committee affirmed (or rejected) the decision. The two exceptions were in Dire Dawa, where the principals of the region's 10 secondary schools catering for grade 10 were involved in the decision-making process and in Somali, where the regional president took a leading role in deciding

how the region's RBA funding would be used. The decision-making processes in both regions were highly context-dependent. The small size of Dire Dawa facilitated an inclusive approach. In Somali a high turnover of REB staff had risked stagnation of RBA activity in the region. Although exceptional, because there was discernible school-level activity in both regions by the end of Year 3, the responses in the two regions highlighted the significance of clear and decisive direction.

For those regions taking the school approach to RBA spending, decision-making processes were devolved to the schools receiving funds. School-based committees working in conjunction with the WEO and PTSA typically made decisions in the context of previously approved school improvement plans.

## 5.2 What Worked and Why

All of the reward payments from the three years of the pilot were fully distributed to 10 of the 11 regions by the end of year 3, and some small changes at the school level occurred by the end of year 3. The remaining reward payments had been designated to fund changes but had not yet been spent when this evaluation was completed. These changes, discussed in Appendix 6, occurred too late to have had any influence on the number of sitters and passers in the EGSECE in EC 2006, but the changes do have the potential to increase the numbers of sitters and passers in the future.

The MoE and REBs appreciated RBA because they had discretionary use of its funding. As a result of the MoE's 'hands off' approach, the regions were able to apply the MoE's guidelines in different ways:

- Replicating the MoE's funding formula for making regional rewards (Dire Dawa, SNNPR);
- Modifying the formula to reward only the most successful schools (Addis Ababa, Oromiya);
- Supporting all schools to increase the numbers of sitters and passers at the regional level (Afar, Harari);
- Supporting schools with clearly identified needs (Amhara, Benishangul-Gumuz, Gambella, Somali); and,
- Supporting underperforming schools to increase the numbers of sitters and passers at the regional level (Tigray).

The first two choices correspond with the school-level approach and the next three with the regional approach as shown in Table 5.1. Each choice reflected exactly what the pilot had intended to achieve. Regional and local officials were given discretion about how to improve educational quality with relatively few restrictions or obligations. Four REBs (Benishangul-Gumuz, Oromiya, Somali, and Tigray) made significant changes to their RBA plans during year 3. These changes suggest the potential for RBA to encourage new ways of thinking about how funding can be used to improve educational quality.

Where the regional approach was taken, interviews with REB officials indicated that autonomy encouraged strategic thinking. They prioritised needs that were most likely to secure greater returns through increased numbers of sitters and passers.

With the school approach, the REBs asked schools to develop their own RBA plans. This meant that some school-level actors had become aware of how RBA works by the end of year 3. Interviews with REB and school staff indicated that the autonomy given to schools also encouraged strategic thinking because the schools also prioritised needs that were most likely to secure greater returns. Nonetheless, this conclusion should be treated with caution given the limited amount of school-level data generated during year 3.

The REBs viewed the freedom to address previously identified problems as a benefit. There was also some limited evidence by the end of year 3 that schools appreciated RBA because it provided some with additional resources. This finding applies, of course, only to those schools that received reward payments and perhaps only to those schools in which the evaluation team conducted interviews.

DFID largely took a 'hands off' approach to RBA throughout the pilot. The MoE initially assumed responsibility for RBA and negotiated with DFID to change the formula and time frame for rewarding the regions. The complaints in year 3 from some regions about the small size of their rewards suggest that the MoE was correct to negotiate these changes and that small initial rewards would have undermined regional engagement with and enthusiasm for the pilot.

The presence of a DFID adviser in the MoE had facilitated these negotiations. Although he did not have any formal responsibility for the pilot, his presence gave the MoE confidence to reflect on the pilot when compared to the MoE's relative inaction following his departure.

The MoE replicated DFID's 'hands off' approach in its dealings with the REBs, but sometimes did not respond to queries from the REBs about the RBA process. This delayed the engagement of some REBs with the pilot. Most of the REBs that complained about the MoE's lack of engagement developed RBA plans corresponding with the regional approach noted above. Although the pattern was not consistent across all 11 regions, it suggests that the MoE's 'hands off' approach enhanced and promoted the REBs' scope for autonomy.

The contrast between the relative inactivity at the MoE and the activity in the regions further suggests that RBA was operating effectively at the subnational level only by the end of year 3 – inasmuch as it was working – because the REBs had directly engaged with it. Most regions were only able and willing to engage directly with the RBA pilot once they had been reassured they would receive their regional rewards, an issue addressed in more detail below.

### 5.3 What Did Not Work and Why Not

As discussed in Chapter 2, there was no evidence of RBA-related changes in the numbers of EGSECE sitters and passers. In addition to the reasons already discussed, the qualitative research identified five possible explanations:

- the misinterpretation of the RBA incentive;
- the lack of sufficient other discretionary funding to pre-fund activities to increase the number of EGSECE sitters and passers;
- the limited amount of RBA rewards to some regions;
- delays in the transfer of the rewards to the regions; and,
- DFID's ambivalence about its roles and responsibilities.

The MoE interpreted RBA as the investment *of* funding rather than investment *for* funding. This interpretation was given to the REBs and, in those regions following the school approach, to schools. Although there was considerable discussion of RBA as an incentive, it was understood as input funding based on results. Interviews with the representatives of the MoE and REBs made clear the assumption that regions and/or schools would use RBA funding to improve EGSECE results and so receive future funding. There was little evidence of changes being made to secure resources from other sources to pre-fund activities that might increase the number of sitters and passers. In other words, potential change was mostly seen as being dependent on the receipt of RBA funds. This is a crucial distinction. As the *Business*

Case observed, 'DFID will pilot an approach where the partner government must arrange finance for the cost of all up-front inputs'. In other words, DFID's expectation was that the prospect of increased financial resources would cause the GoE to use *existing resources* to increase the number of sitters and passers and only then be rewarded for the results achieved. This may have been a questionable expectation. As Sarah Vaughan noted in her assessment of the Ethiopian pilot, the 'GoE strongly discourages the pre-financing of non-funded expenditures, and the system is an input-based one....GoE expenditures are authorised on the basis of an approved and funded budget'<sup>51</sup>.

The MoE also interpreted the incentive aspect of RBA as competitiveness. It explained that REBs would be encouraged to increase the quality of education in the regions to secure more funding and more funding than other regions. Some REBs put forward similar arguments about interzonal, district, and/or school competition.

Once the mechanism for transferring RBA funds to the regions had been agreed, the MoE advised REBs to use other, non-RBA funds as an interim measure to finance RBA-related changes. The only REB to do this was in Gambella. The other REBs explained that they did not have such funds available or were apprehensive that they would not be reimbursed even if they did have access to non-RBA funds. The evaluation team's baseline and interim evaluation reports highlight the disparity between needs and funding in many Ethiopian schools. The reports reveal the ingenuity of some schools in raising funds but when it is realised that they are often spent on basics – such as chalk, textbooks, and uniforms – they also indicate the severe impoverishment of those schools and the scale of funding needed to make meaningful changes to the quality of education. This inevitably calls into question the efficacy of other funding mechanisms (e.g., GEQIP), but it also highlights the lack of sufficient discretionary funding to finance change prior to receiving the reward payments associated with the pilot.

The MoE's interpretation of the RBA incentive has the potential to encourage change: once funding is received, REBs can invest it to secure returns in the improved EGSECE results that will generate future funding. Nonetheless, the amount of RBA funding is small compared to other funding mechanisms such as GEQIP<sup>52</sup>. REBs and schools in receipt of relatively small rewards noted that these amounts were insufficient to make any significant changes. This was a particular concern in the emerging regions. As an illustration, Afar's REB chose to supplement existing teacher training programmes because the small size of its reward did not allow it to make any other viable investments such as providing additional much-needed materials for schools. In any event, the REBs did not receive their rewards until or after the end of the EC 2006 school year and so there was no time to use the rewards to improve EGSECE results by the end of the pilot.

#### 5.4 Targeted Funding

The importance of investing in girls' education is widely recognised at all levels – from the MoE through the REBs to schools. The actors at these levels recognised the potential for RBA to support girls' education as they engaged with the pilot. They also acknowledged the RBA premium payable for increased numbers of female sitters and passers. Despite this acknowledgement, few reward payments were invested in female

<sup>51</sup> Sarah Vaughan, 'Going against the Grain? Lessons from a DFID Results-Based Aid pilot in Ethiopia', Policy brief completed for DFID/Ethiopia, February 2015. DFID's 2014 annual review of the pilot makes the same point: 'requiring REBs/districts/schools to plan and implement activities before they have funds is against GoE budget guidelines and undermines the public financial management and fiscal discipline, which the GoE and donors have been trying to instil over the years'. See DFID, 'Annual Review: Pilot Project of Results Based Aid (RBA) in the Education Sector in Ethiopia', December 2014. Available at [iati.dfid.gov.uk/iati\\_documents/4839826.docx](http://iati.dfid.gov.uk/iati_documents/4839826.docx)

<sup>52</sup> GEQIP II (2014-18), a multi-donor, pooled education programme, has a projected budget of \$550 million, including approximately \$186 million from DFID.

students, at least among the schools visited for the evaluation. Examples include the formal recognition of their achievements in Gambella, the proposal to construct a separate toilet block at a school in Oromiya, the construction of a female-only cafeteria at a school in Somali, and the provision of additional resources for girls across the region. The main focus in all regions was on improving the overall conditions for learning. When questioned, representatives of the REBs and schools explained that girls would benefit from this focus. This approach does not increase the disparities between boys and girls but does little to reduce it.

The REBs of the four emerging regions noted that their RBA rewards would make only a small difference in the wider context of the considerable difficulties they face in increasing the quality of education. Their concerns were emphasised by the regional rewards distributed over the pilot's three years. Six regions, including three of the four emerging regions, received lower reward payments than they had earned based on their increases in the number of sitters and passers, even when considering the amounts that the MoE retained (see Appendix 2). This suggests that the situation confirms the discussion in Chapter 3, namely that the RBA pilot has not done much to address the disparities between the emerging and non-emerging regions.

Finally, there appears to have been uncertainty and some ambivalence within DFID about what its role should have been once the pilot started. The MoU was largely silent about DFID's role in implementing the pilot. DFID consequently assumed that the GoE would take full responsibility for the pilot, but the MoE struggled to adapt to the new aid format.

The presence of the DFID adviser in the MoE undermined the assumption that key decisions could be left to the MoE, at least in the pilot's first year. The adviser was instrumental in the completion of the MoE's guidelines, but the pilot largely stalled within the ministry after his departure. This situation may have created a dilemma for DFID. Intervening to prompt the MoE's action would have violated one of the ideas being tested in the *Business Case* and jeopardised what was being evaluated. In contrast, adhering to the 'hands-off' approach risked the pilot's success, but this seems to be a risk that DFID reasonably assumed in the interest of testing the pilot's implementation. By the end of the pilot, however, most of the REBs had assumed considerable responsibility for the pilot and for decisions about how to spend the rewards they received.

## 5.5 Conclusions

Although the pilot did not accomplish as much as had been hoped, based on the amount of available rewards, there is evidence of RBA's potential to improve the quality of education in the future. RBA was being used to fund some regional and school-level changes during the pilot's last year, and more RBA-funded changes were planned for the following year. Furthermore, the strategic thinking of REBs taking the regional approach and schools in those regions taking the school approach indicates their willingness and ability to target specific needs.

In most regions the 'trickle down' effect followed the receipt of RBA funds. That is, most REBs engaged with RBA only when their rewards had been confirmed. All the REBs focused on using their RBA rewards to meet previously identified needs.

The original three-year period for the pilot was too short, as DFID has acknowledged<sup>53</sup>. More time has been required than envisaged to establish the initiative and to operate it. After the pilot's unexpected delays, however, the RBA system had finally begun taking effect at the regional and school level at the end of year 3. Interviews with representative of the REBs and schools indicate that it is not unreasonable to suggest that the main incentive was the discretionary nature of the rewards rather than their size. Now that rewards are being spent there is potential for future evaluations to consider the efficacy and effectiveness of the different approaches to the use of RBA.

---

<sup>53</sup> DFID/Ethiopia, 'Terms of Reference (TOR) for Contracting an individual consultant for Independent Verification of Key Government of Ethiopia (GoE) Educational Data of 2014/15 for a Pilot of Results-Based Aid (RBA) – Ethiopia', 2015.



## 6 Impact on Aid Relationships

DFID is the single largest donor in Ethiopia's education sector, enjoying a special level of policy dialogue with the government and with the MoE. DFID is attracted to the RBA model because it focuses dialogue on the achievement of agreed outcomes. The MoE has shown genuine interest in funding mechanisms that reduce conditionalities, an objective consistent with the Paris Declaration on Aid Effectiveness. Other donors have expressed interest in the potential of RBA as an aid modality, partly reflecting DFID's advocacy on the global stage in promoting RBA and being at the forefront of efforts to increase the use of RBA in Ethiopia and elsewhere.

DFID's aspiration that RBA might improve the dialogue between DFID and the MoE has not, however, yet been realised for several reasons. The enthusiasm expressed in principle by the MoE for RBA has been offset by several concerns about the nature of the pilot itself. As detailed in the team's interim evaluation report, these concerns include the relatively small amounts of money available through RBA and a perception of high transaction costs. These costs would include: (a) negotiations with MoFED regarding the flow of funds, and (b) work involved in creating RBA-related structures for national-to-regional-to-school disbursements. Much of the dialogue between the MoE, DFID, and other donors understandably revolves around implementing new phases of PBS and GEQIP. These projects represent sizeable investments in the education sector. The MoE has also questioned mechanisms used for RBA in that they could lead to an inequitable distribution of rewards among the regions.

As noted in Chapter 1 (and Appendix 2) this questioning led to a modification of the formula used to reward regions once the results from the EC 2004 EGSECE had been announced. In other words, *there has been a preoccupation in dialogue between DFID and the MoE with overall amounts of money and how RBA rewards ought to be distributed rather than with the core purpose of RBA, namely on how these rewards can be used to encourage improved outcomes*. On DFID's side there has also been frustration at the perceived lack of leadership at the national level in communicating a clear vision to the regions on how RBA can be used. The problems in implementing the RBA pilot may have been impaired by DFID's 'hands off' approach in its dealings with the MoE, which was consistent with the ideas included in the *Business Case* and the subsequent MoU. Similarly, both parties may not have appreciated the time and effort required to achieve behavioural change in relation to new aid modalities.

Interviews with DFID officials reveal recognition that there has not been the anticipated behavioural change on the part of the MoE with regard to RBA. DFID is also aware of many of the problems associated with the pilot's design. For example, DFID officials suggested that future attempts at implementing RBA in the education sector should be based on indicators that measure changes in added value and improvements in raw scores in examinations. In the short term, and in relation to the existing pilot, there is a realisation of the need for more structured engagement between the MoE and the MoFED in implementing RBA. Likewise, regular meetings between DFID and MoE are desirable and should aim to facilitate implementation and make it more efficient by putting in place clear lines of communication. Encouraging the MoE to take a proactive leadership stance in relation to the regions and how RBA money ought to be spent would also be desirable in the eyes of those interviewed.

There is also evidence of a degree of adaptation at the national level although these are mainly reflective of the MoE's attempts to increase the overall amount of the RBA rewards. The negotiated changes in the allocation of rewards to the regions provides an example. This is, however, more of an adaptation to consider the regions' political economy rather than to reward improvement. The use of RBA at a regional level suggests that the REBs have used some of their RBA rewards to plug existing funding gaps (see Appendix 6). It would be interesting to see whether the use of RBA rewards at a regional level will lead to a

more strategic targeting of resources and the more even development of capacity within regions on which future improvements in outcomes can be built.

As DFID's *Business Case* notes, the 'pilot is also expected to impact on other donors' institutional approach to the delivery of aid'. As DFID further commented, 'a measure of the pilot's overall success will be the extent to which other stakeholders, including GoE and other donors, perceive RBA as being an effective instrument and whether it leads to similar arrangements either in the education sector or beyond'. The evidence suggests that these objectives have not yet been achieved.

There has been a lack of communication at the national level between DFID and other donors about RBA generally and the pilot in particular. This is in contrast to the advocacy role that DFID has taken in relation to RBA on a global scale. Thus, although all of the donor representatives interviewed expressed interest in the RBA pilot, there was uneven awareness of the pilot among these representatives. In addition, there has been limited communication in meetings among donors about the results of the pilot. This situation may reflect the limited impact that RBA has achieved. While all of the donor representatives interviewed remained positive about the potential of RBA as an aid modality, this support was qualified by a recognition of some of the potential challenges associated with RBA.

Although these qualifications were not expressed with direct reference to the RBA pilot, they speak to some of the design and implementation issues encountered in the pilot. For example, interviews with staff of the World Bank and the U.S. Agency for International Development highlighted the importance of linking rewards for improved performance to outcomes in areas in which there exists demonstrable capacity across the sector to deliver: otherwise only historically high performing parts of the system would be rewarded. *The implication is the need to front load and target funding to ensure the capacity to institute change before implementing RBA.*

Given the events that transpired during the pilot, what can be concluded about its impacts on relationships? Have there been significant and desirable changes in the DFID-MoE dialogue as a consequence of the pilot? Initially, at the stage of design of the pilot, there was intensive dialogue involving DFID's senior education adviser embedded in the MoE's Planning Directorate. After his departure, successor DFID advisers have faithfully followed the dictum of taking a 'hands-off' stance in relation to communications with MoE over the progress of the pilot, restricting their interactions to discussions about the delivery of rewards and adjustments to the original levels of rewards. Only in October 2013, notably at the MoE's annual education conference, when it became increasingly apparent that little activity was being initiated, did DFID advisers begin to take a more proactive role, by insisting that the minister of education should speak to an agenda item relating to the RBA rewards.

## 7 Value for Money

This chapter considers whether the benefits of RBA outweigh its costs. In doing so it is important to distinguish between *the RBA instrument* and the *actual use of RBA payments*. The focus of this evaluation is primarily on the former, but the evaluation team was also asked to consider whether RBA is a more effective instrument than two other DFID-supported programmes in Ethiopia, Promoting Basic Services (PBS) and the General Education Quality Improvement Programme (GEQIP). DFID also expressed an interest in knowing whether the rewards were set at an optimal level. The approach used here has benefitted from consideration of the methods and approaches in use by the DFID-funded team conducting an evaluation of the department's RBA pilot in Rwanda.

### 7.1 Methodology

The analysis attempts to identify and, to the degree possible, quantify and monetise, the costs and benefits associated with the RBA pilot. The analysis also tracks progress against a range of indicators related to value for money (VfM). A detailed discussion of the methodology is in Appendix 7. Readers should appreciate that assessing value for money is not an exact science, and different approaches have strengths and weaknesses. Such weaknesses can be mitigated by making plausible assumptions, considering several scenarios, and by being conservative in drawing conclusions, as the following discussion attempts to do.

As an illustration, the analysis considers only those benefits that are reasonably *attributable* to the RBA pilot. The incremental benefits of the RBA instrument itself include the economic returns from having more males and females passing through the education system (and ultimately earning higher incomes and enjoying better health) but also less tangible benefits such as enhanced female empowerment. The benefits also include any effects on system efficiency directly attributable to the RBA. The incremental costs associated with the RBA instrument include the costs of verification as well as the costs of educating any additional students sitting and passing that are attributable to RBA<sup>54</sup>.

Similarly, the analysis considers the incremental costs and benefits associated with the use of the RBA reward payments. Costs include the value of the RBA disbursements to the MoE. The benefits consist of the impact this spending might have<sup>55</sup>. Of particular interest here is the question of whether RBA payments would have been better spent on more appropriate items or activities than PBS or GEQIP. These issues are addressed in Appendix 6, which discusses how the RBA funds were allocated, and through analysis of available studies on the impacts of the PBS and GEQIP.

The evaluation team expressed concerns at the outset of the pilot that the decision to use a rolling baseline as a basis for reward payments created the possibility that DFID might simply provide rewards based solely on the counterfactual (i.e., results that would have been achieved without the pilot)<sup>56</sup>. To assess the implications of this possibility, the analysis further considers several different counterfactuals to estimate what payments might have been made had alternative baselines been used. These include the

<sup>54</sup> The counterfactual for the RBA pilot is a 'do nothing' option. DFID made it clear that the RBA funds would not otherwise have come to the education sector.

<sup>55</sup> With RBA the government has full discretion over use of RBA funds. A country that achieves the desired outcomes and triggers reward payments but allocates none of the rewards to the sector responsible for achieving the outcomes (and therefore shows no additional sector benefits) is just as successful as one that allocates all of the reward to the sector (as is the case for the Ethiopia pilot). In evaluating other RBA programmes one would need to be careful not to discriminate against countries that allocate rewards outside the sector that produced the desired outcomes.

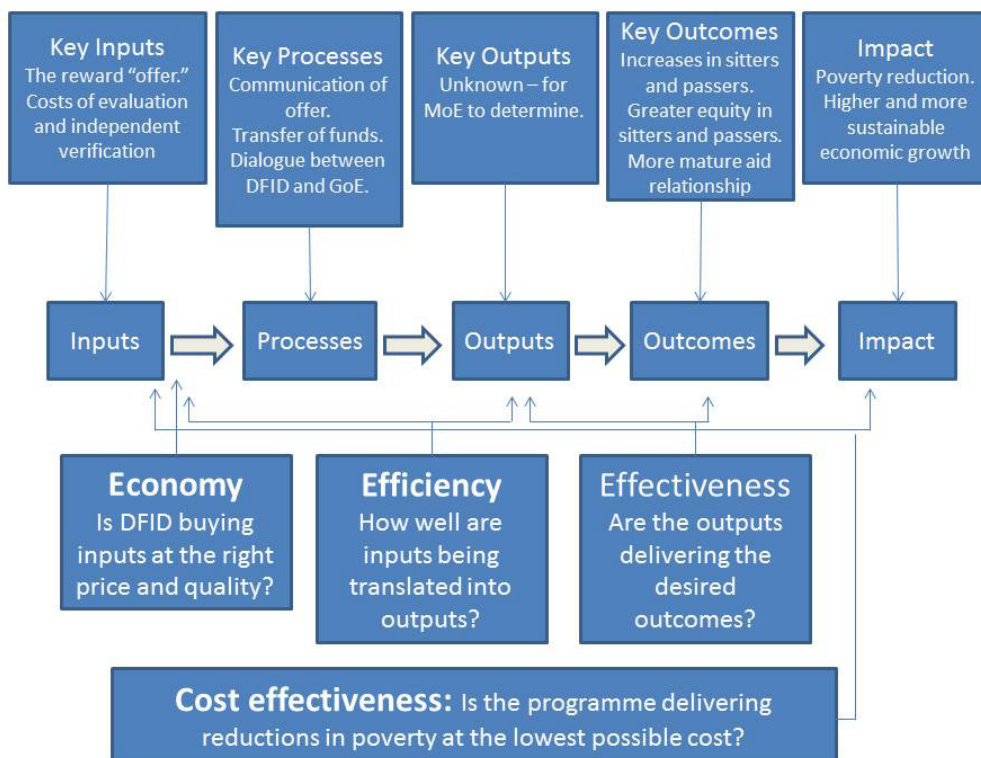
<sup>56</sup> With a rolling baseline, as noted in Chapter 1, the base year is adjusted annually with performance one year used as the baseline for the next year. For example, EC 2003 was used as the baseline year for results in EC 2004 and that year was used as the baseline for results achieved in EC 2005. Ironically, DFID's *Business Case* asserted that the rolling baseline would minimise the risk that the project would pay for results that would have been achieved anyway.

counterfactual that DFID estimated during the design process as well as ones the evaluation team developed, such as adjusting according to previous trends (EC 1999 to EC 2003), likely growth in the overall pool of potential grade 10 enrolees, and according to enrolment rates in previous years<sup>57</sup>. The effects of using these different counterfactuals on the level and timing of reward payments and on the source of the reward payments are also assessed<sup>58</sup>.

An additional VfM metric of cost to DFID per additional *net* sitter and passer is considered to allow for the fact that while the GoE received payments for regions that had increased the numbers of sitters and passers there were no adverse consequences for regions in which the numbers of sitters and passers declined<sup>59</sup>. The cost per additional net sitter is compared to the actual rewards made and to the unit costs of delivering secondary education to grade 10.

DFID’s approach to VfM has been used, recognising that the point of RBA is that while the relation between inputs (the RBA rewards) and outcomes (grade 10 sitters and passers) is defined in advance, the outputs required to achieve them are not. This is illustrated in Figure 7.1, which is adapted from DFID’s guide to VfM and which identifies some of the key VfM questions<sup>60</sup>.

Figure 7.1: Value for Money Approach



Source: DFID, 'DFID's Approach to Value for Money (VfM)'.

<sup>57</sup> For year 1, adjustments were made according to the growth in grade 10 enrolees. For year 2, adjustments were made by growth in grade 10 enrolees that year and grade 9 enrolees the previous year. For year 3 adjustments were made for growth in grade 10 enrolees, growth in grade 9 enrolees the previous year, and grade 8 enrolees the year before that.

<sup>58</sup> This refers to the sources for the payments made to the GoE and not the subsequent allocation of the payments to the regions.

<sup>59</sup> For simplicity sake sitters and passers are weighted equally and the results are aggregated. Separating them does not add much to the analysis.

<sup>60</sup> DFID, 'DFID's Approach to Value for Money (VfM)', July 2011. Available at [https://www.gov.uk/government/uploads/system/uploads/attachment\\_data/file/67479/DFID-approach-value-money.pdf](https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/67479/DFID-approach-value-money.pdf)

In terms of assessing the design of the payment structure the team took as its starting point the proposition that VfM would be maximised if the rewards were (a) set at the optimal level and (b) focused only on additional results that could be attributable to RBA. Doing so corresponds with DFID's guidance on payment by results.

## 7.2 Cost Effectiveness of the RBA Instrument

The econometric model discussed in Chapter 2 did not detect any statistically significant effects of the pilot on passers and sitters. As a consequence, it is not possible to conclude that the RBA instrument has provided value for money. The net present value (NPV) of the base case, which reflects the cost of the independent verification, is estimated to be -£0.31 million using a discount rate of 12 per cent<sup>61</sup>. To breakeven, the pilot would have had to increase the number of sitters by about 300<sup>62</sup>. This suggests that RBA as an instrument is a low-cost, relatively low-risk approach because its costs are low if it does not work but with potentially high rewards albeit with higher costs if it does work. Put another way, the pilot would have had to have achieved few additional attributable sitters to have broken even. This, in itself, presents a strong justification for DFID having initiated the pilot<sup>63</sup>.

## 7.3 Relative Cost Effectiveness of RBA with PBS and GEQIP

In the absence of an ability to attribute any increases in the number of sitters and passers to the RBA pilot, it is appropriate to conclude that if GEQIP and PBS achieve any impact they will have been more cost effective than the RBA pilot. Nonetheless, had an RBA impact been observed it would still not have been possible to compare the relative cost effectiveness of the three programmes. The available studies do not address cost effectiveness in a comparable way. The evaluation of GEQIP was qualitative and reported key stakeholders' perceptions of the programme's effectiveness<sup>64</sup>. The evaluation of the PBS was more methodologically rigorous but did not provide comparable measures of cost effectiveness<sup>65</sup>. In the future, if DFID wishes to compare the relative cost effectiveness of its interventions it should ensure the use of common measures and methods across the interventions.

An analysis of the planned use of RBA funds (in Appendix 6), suggests that if these plans are implemented as proposed these funds can be used just as effectively as GEQIP or PBS funds. Although one might expect and hope that the money disbursed to the regions and their schools will be used well, readers should appreciate the appropriate distinction between disbursements, planned expenditures, and actual expenditures. The evaluation team had complete information on the first issue, partial information on the second, and limited information on actual expenditures because the data collection for the evaluation was completed before all the rewards payments had been disbursed and spent<sup>66</sup>. Similarly, no data are available on the effects of the expenditures, and one should not expect that these intended effects will be immediately observable in part because of late disbursements to the regions and even later disbursements to schools. It remains to be seen, therefore, what the impact of this spending will be, and the evaluation

<sup>61</sup> Net present value is the sum of benefits and costs each discounted at an appropriate discount rate. DFID's *Business Case* used a discount rate of 12 per cent. No schooling costs are applicable because the model detects no attributable impact.

<sup>62</sup> This assumes 50 per cent males and 50 per cent females. Only 145 additional female sitters would be needed to breakeven.

<sup>63</sup> Given that the number of additional attributable sitters required for breakeven is within the confidence limits of the econometric model discussed in Chapter 2 one cannot categorically conclude that the pilot has not achieved value for money. In other terms, the pilot's design did not provide the econometric model with the power to detect such small impacts.

<sup>64</sup> HIFAB International AB, 'Comprehensive Evaluation of the GEQIP Programme – Final Exit Survey', June 2013.

<sup>65</sup> ITAD, 'Value for Money Assessment of the Protection of Basic Services Programme', February 2015.

<sup>66</sup> Some information on expenditures at the regional and school level through late 2014 is provided in DFID's 2014 annual review of the pilot. See DFID, 'Annual Review: Pilot Project of Results-Based Aid (RBA) in the Education Sector in Ethiopia'. Available at [iatf.dfid.gov.uk/iati\\_documents/4839826.docx](http://iatf.dfid.gov.uk/iati_documents/4839826.docx)



team declines to speculate on what these future impacts will be. Nonetheless, given that the use of RBA funds is subject to fewer safeguards than with PBS and GEQIP, the risks of not achieving the desired impacts may be higher. A key question for DFID is whether it would have approved a proposal to spend £15.6 million on the basis of the existing MoU as a stand-alone project.

#### 7.4 Optimal Level of Incentives

Judging whether rewards were set at an optimal level would require a comparison of different reward amounts. This would mean, for example, a situation in which some regions (or, preferably, districts or schools within the same regions) would be offered higher (or lower) rewards per sitter or passer than in other districts or schools in the same regions. As already noted, however, DFID chose not to vary the size of the incentives within regions. Although the pilot did offer different amounts for additional sitters and passers in emerging versus non-emerging regions, the two sets of regions are so distinctly different that any comparisons of the reward structure would be invalid. For the same reason using the impact on males and females as a means of assessing the optimal level of reward is also invalid.

#### 7.5 Effect of Using Alternative Baselines

During the pilot's design DFID modelled the likely impacts of the RBA pilot and a counterfactual and then included the results in the *Business Case*. The modelling projected that the increase in the number of male and female sitters would be at least two times higher with the pilot than without it. Similarly, the modelling projected that the number of female passers in the emerging and non-emerging regions would be at least 75 per cent higher with the pilot compared with the counterfactual, namely the situation without the pilot.

The *Business Case* also identified, as a high risk, the possibility that DFID would inappropriately pay rewards for results that would have been achieved in the absence of the pilot. The *Business Case* observed that DFID's modelling of the results and the use of an adjusting baseline were expected to minimise this risk, although it is not clear why such a baseline would minimise the risk. In the end, DFID chose what it considered to be a simple, transparent, and easily understood baseline, namely the number of sitters and passers in the prior year.

From DFID's perspective this baseline had several advantages. On the one hand, information on the annual number of sitters and passers is readily available and easily determined. Students register in advance for the examination, and the number of students actually sitting for and passing the examination is also known and publicly reported in the MoE's *Education Statistics Annual Abstracts*. On the other hand, this baseline does not require estimates or projections of trends in enrolment, such as the number of students finishing grade 9, the number then proceeding to grade 10, and the number of these grade 10 students who finish the academic year and then decide to register and actually sit for the EGSECE for each of 44 groups of students -- girl sitters, girl passers, boy sitters, and boy passers for each of the 11 regions. Moreover, there is no reason to believe that the trends in progression from grade 9, completion of grade 10, and sitting for and passing the EGSECE for each of the 44 groups would be remain constant from one year to the next.

In short, given the choice between transparency and administrative efficiency versus potentially lower reward payments based on what might have been more complex baselines, DFID opted for the former. Nonetheless, it is still of value to consider the potential financial consequences of different baselines. One of DFID's objectives for the evaluation is to inform other donors about RBA, and their preferences for baselines may differ from DFID's. The evaluation team thus compared the VfM using the counterfactual that DFID identified during the design phase (but not ultimately used in the final design) and eight

alternative scenarios: grade 10 population pool, past trends (EC 1999 to EC 2003), and six based on different combinations of assumptions about adjustments for years 1, 2, and 3. In the absence of disaggregated data on enrolment rates for EC 2006 it was not possible to assess two of the scenarios, namely those adjusting according to grade 10 enrolments in EC 2006. These alternatives are discussed in greater detail in Appendix 7.

### Using the Design Stage Counterfactual

Paying rewards using a rolling baseline rather than only on the numbers of sitters and passers above DFID's modelled baseline resulted in a net difference of approximately £2.6 million, which means that DFID would have provided approximately £13 million in reward payments rather than the £15.6 it actually provided. *With the latter amount DFID provided payments (i.e., £2.6 million) for increases that likely would have occurred in the absence of the pilot.* One implication is that rather than spreading the rewards over all additional sitters and passers above the rolling baseline, as was the case with the pilot, DFID could have focused the rewards only on those additional sitters and passers estimated to be attributable to RBA. Doing so could have sharpened the incentives through providing a higher reward for each additional sitter and passer.

### Using Alternative Counterfactuals

Figure 7.2 shows that had alternative baselines based on the counterfactuals the evaluation team developed:

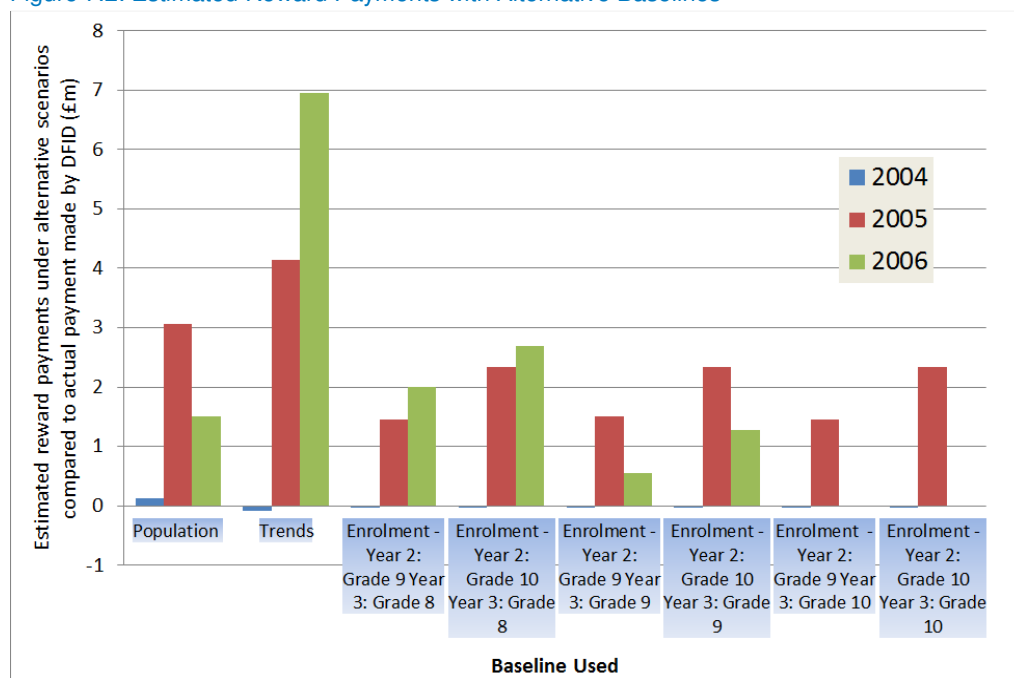
- DFID would have paid less overall than it actually did in almost every scenario in every year;
- payments would have had a different time frame with smaller payments in years 2 and 3; and,
- payments would have been generated from different regions<sup>67</sup>. This implies that the government received reward payments on the basis of performance in some regions but did not receive payments on the basis of performance that – based on the revised baseline – was actually much better.

---

<sup>67</sup> This is illustrated in Appendix 7.



Figure 7.2: Estimated Reward Payments with Alternative Baselines



Note: Disaggregated data on enrolments were not available, so it was not possible to estimate the amounts for EC 2006 for the two scenarios on the far right. The figure does not include DFID’s modelled baseline included in its Business Case.

As an illustration, had the baseline been adjusted according to the growth in the size of the grade 10 pool (the scenario on the far left of Figure 7.2), DFID would have provided £0.13 million, £3.07 million, and £1.50 million respectively for results achieved in years 1, 2, and 3 of the pilot.

Under all of the alternative baselines DFID would have paid less in years 2 and 3 than it actually did. This is particularly true for the scenario in which the baseline was increased year on year in the ‘trends scenario’. Under such a baseline the difference in reward payments over the three years would have amounted to about £11.1 million. Among the eight scenarios considered in Figure 7.2 the net differences between actual reward payments and those associated the alternative baselines ranged from £2 million to £11 million.

Depending on one’s objectives, these amounts may be considered unduly large or tolerably modest for an experimental project designed to test a new aid modality. Providing £11.1 million less than was actually provided might be appealing to the advocates of increased VfM who believe that payments should be made only on the basis of performance attributable to the pilot. One can then ask, however, whether the ‘meagre’ rewards thus provided would have been sufficient to motivate or encourage any meaningful changes at the national, regional, or school level. DFID’s objective with the pilot was not to minimize its costs but rather to test the theory of change. The reward payments for each additional sitter and passer that DFID chose were believed to be sufficiently enticing to encourage the desired changes – without knowing in advance whether they were set at the proper level. If DFID had wanted to minimise costs it could have offered lower reward payments for increases in the number of sitters and passers, but doing so might have eliminated or eviscerated the assumed incentives associated with RBA.

Analysis of the regional breakdown of the rewards also reveals that the use of different baselines would have meant that rewards payments would have been generated differently. The alternative baselines

suggest that under the current design, including the MoE's adjustments, the performances of Addis Ababa, Dire Dawa, and Harari were 'underrewarded' while those of SNNPR, Amhara, and Tigray were relatively 'overrewarded'.

## 7.6 Cost per Additional Net Sitter

As shown in Table 7.1 the net cost to DFID per additional sitter/passers declined from £1,238 at the end of year 2 to £123 by the end of third year<sup>68</sup>. The year 2 figure is high because at the end of that year DFID had provided over £6.5 million yet the numbers of sitters and passers had only exceeded the baseline levels following the reductions in year 1. Although year 3 saw a large payment by DFID, the number of net additional sitters and passers compared to the baseline increased.

Table 7.1: Cost per Additional Net Sitter/Passer

	At end EC 2004	At end EC 2005	At end EC 2006
Annual payment (£)	896,260	5,664,935	8,995,215
Cumulative reward payments (£)	896,260	6,561,195	15,556,410
Net increase in sitters + passers (compared to EC 2003)	- 68,707	5,301	126,226
Net cost (£) to DFID per additional sitter and passer	n/a	1,238	123

The net cost per additional sitter/passers of £123 compares to an estimated average cost of funding two years of secondary education of £50 (*Business Case*) and an average reward payment per additional sitter and passer of £71.8 and £71.6, respectively<sup>69</sup>. This suggests that the current reward structure provides a reasonable financial incentive to increase passers and sitters because it far exceeds the average costs to the education sector of generating these passers and sitters.

## 7.7 Comparisons with Assumptions in DFID's *Business Case*

As shown in Table 7.2 modelling conducted for the *Business Case* assumed 191,047 additional sitters and 178,021 passers<sup>70</sup>. In contrast, the additional results achieved are well below those anticipated in the base case (although this is not the case for male and female sitters and passers in emerging regions where results exceeded those projected in the *Business Case*) and attributable results (as shown in Chapter 2) are zero.

<sup>68</sup> Sitters and passers are weighted equally.

<sup>69</sup> The evaluation team did not have access to information on actual expenditures for secondary education, so the £50 estimate cannot be updated.

<sup>70</sup> These figures for sitters and passers are based on the spreadsheets used for the *Business Case*. Figures were rounded up or down in the *Business Case*.

Table 7.2: Comparison of RBA Pilot Results with *Business Case Assumptions*

	Business Case		Actual Results	
	Additional Results	Attributable Results	Additional Results*	Attributable Results
Non-emerging sitters: male	55,127	38,077	24,382	0
Non-emerging sitters: female	129,065	68,358	62,414	0
Emerging sitters: male	3,267	2,182	7,057	0
Emerging sitters: female	3,588	1,997	4,363	0
Non-emerging passers: male	70,391	21,597	2,437	0
Non-emerging passers: female	100,437	48,090	18,261	0
Emerging passers: male	4,557	1,681	5,068	0
Emerging passers: female	2,636	1,310	2,244	0
Overall: sitters	191,047	110,614	98,216	0
Overall: passers	178,021	72,678	28,010	0
Overall: male	133,342	63,537	38,944	0
Overall: female	235,726	119,755	87,282	0

Based on data from the independent verification reports.

Source: Original spreadsheets supporting DFID's *Business Case*.

The *Business Case* projected the costs of educating pupils to grade 10 at £73 and values this at £87.6 (based on a 20 per cent return to secondary education). Using these figures and the assumption that RBA will increase sector efficiency by 0.6 per cent per annum, DFID projected a NPV of £50.25 million declining to £36.1 million under an alternative scenario assuming lower improvements in efficiency.

## 8 Potential Unintended Consequences of the RBA Pilot

Soon after the pilot began the evaluation team identified several potential unintended consequences, shown in Table 8.1, for the education system that *could* result from the RBA pilot. In addition, the table also shows the likelihood of the consequence occurring and the impact or seriousness if it were to occur.

Table 8.1: Potential Unintended Consequences of the RBA Pilot

	Potential unintended consequence	Likelihood	Potential impact
1	Diversion of efforts from other sectors or other sub-sectors in education	L	M
2	Additional funds in the system might increase pressures for corruption	M	H
3	Pass rates increase due to structural changes in EGSECE	L	H
4	Schools reduce grade-repetition rates	L	L
5	More cheating by schools	M	H
6	Schools focus attention on students near EGSECE pass/fail threshold at expense of others	M	L
7	Teachers move to successful schools in search of additional financial rewards	L	L
8	Children from poor households and pastoralist communities fall further behind others	M	M

L = Low; M = Medium; H = High

Before discussing each of these possible consequences, two issues merit attention. First, not all of the consequences are necessarily negative (as 'unintended' is often taken to mean). Some unintended consequences could be neutral or positive. For example, re-allocation across sectors and sub-sectors or reducing grade repetition might not be negative. These possible consequences are considered because they were not identified as planned consequences in DFID's *Business Case*. Second, given the limited response to the RBA pilot noted in previous chapters and the lack of compelling evidence for the impact on the pilot's *intended* consequences, it may be likewise difficult to find evidence of *unintended* consequences in the same period.

To begin, and as discussed in the chapter on system effects, no increases were detected in the relative emphasis on lower secondary school at the expense of other levels, sectors, or sub-sectors in education (potential consequence #1), and neither was the evaluation team able to detect increasing pressures for corruption (potential consequence # 2).

Changes in the pass rate affect the numbers passing the EGSECE and, hence, the rewards under the pilot (potential consequence #3). The EGSECE pass rate in the three pilot years and five preceding years are shown in Table 8.2.

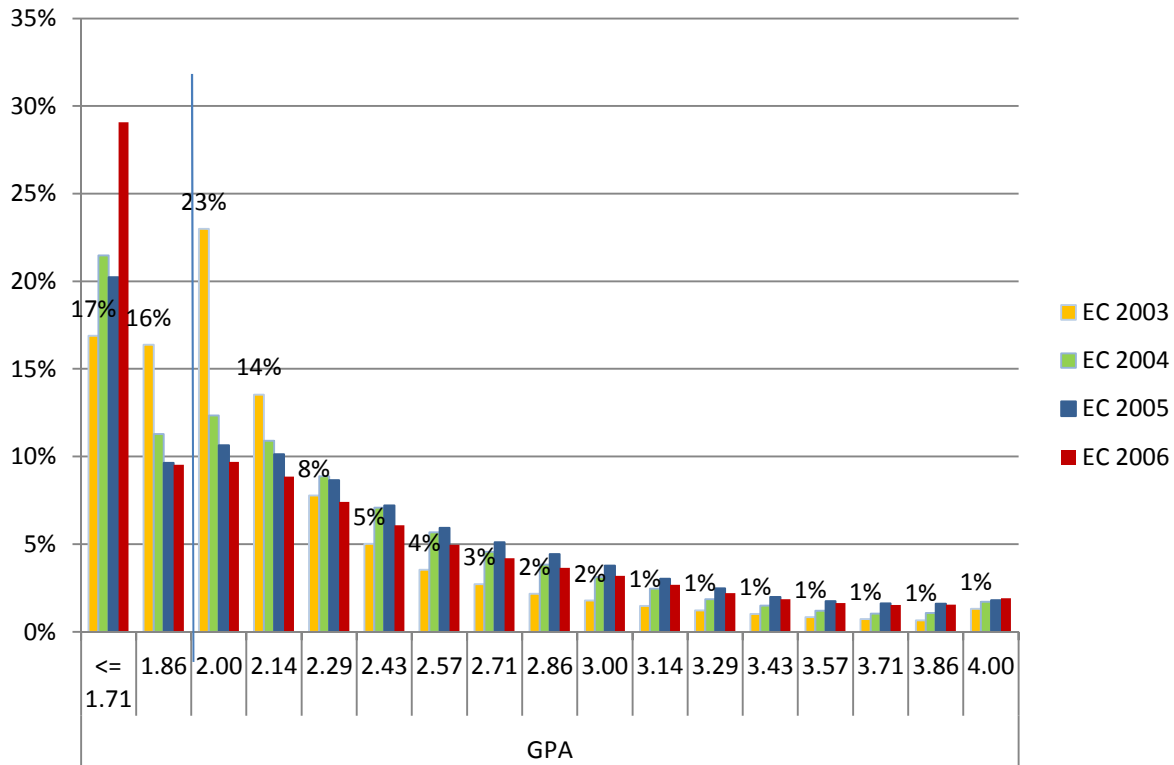
Table 8.2: EGSECE Pass Rates, EC 1999-2006

EGSECE Pass Rates, EC 1999-2006								
Year	EC 1999	EC 2000	EC 2001	EC 2002	EC 2003	EC 2004	EC 2005	EC 2006
Pass rate (%)	49.8	38.4	42.6	62.3	66.5	67.2	69.4	60.9

As the Coffey verification team reported, 'The fluctuating pass rates prior to [EC] 2002 indicate systemic changes to the examination system'. No systemic changes have been reported since EC 2002. The pass rate has been more consistent than before but arguably not as consistent as the stated norm-referenced

nature of the individual subject tests would indicate<sup>71</sup>. The EGSECE pass rate may be susceptible to relatively small changes in the difficulty of individual subject tests and the application of an occasionally used alternative threshold between a D grade and a failure in the subject. Figure 8.1 charts the distribution of EGSECE sitters' grade point averages (GPA) at the level of individual subject grades.

Figure 8.1: GPA Scores in EC 2003, 2004, 2005, and 2006



Two changes occurred in the percentage of students who passed the EGSECE. First, in EC 2003 the pass/fail status of nearly 40 per cent of sitters could have been altered by a change of only one grade in a single subject (e.g., to a B from a C). This percentage was roughly half as large in the following three years: a positive development, though one that might have led to a large change in the pass rate. Second, the number of sitters obtaining a GPA of 1.71 or less jumped by nearly ten percentage points in EC 2006, and the proportion of sitters missing a pass by the equivalent on one grade on six or more subjects increased to 4 per cent from 1 per cent.

Neither of these changes is likely to reflect the learning achievement of the students sitting for the test, so they suggest the examination's administration changed, though NEAEA did not advise the verification team or evaluation team of any changes. Neither change would affect the ranking of the highest achieving students for selection for upper secondary education. Most important, there is no evidence that these changes were linked to the pilot.

The evaluation team estimates the reward payment in EC 2005 to have been £1.2 million higher than it would have been had the pass rate been the same as in EC 2003<sup>72</sup>. The decrease of nearly nine percentage points (to 60.9 per cent in EC 2006 from 69.4 per cent in EC 2005) will make it easier to obtain

<sup>71</sup> In its *Baseline Report*, the Coffey verification team stated that the EGSECE pass rate from EC 2002 onwards should lie in the range 60 to 65 per cent, while in its *Phase 3* and *Phase 4 reports* the verification team suggested that the pass rate should be in the range 67 to 70 per cent. These ranges do not overlap, and both ranges miss three of the five values from EC 2002 onwards.

<sup>72</sup> Calculated by 2.9 per cent x 550,000 sitters x £75 [an average of payments between £50 and £100 depending on sex and region].

rewards for passers if the RBA pilot is continued in EC 2007 (because the baseline for the latter year will be EC 2006).

Neither the evaluation team nor the verification team found evidence that the examination's administration has been altered in any way to increase payments (potential consequence #3). The NEAEA does not receive any RBA payments, and the MoE disbursed 97 per cent of the RBA payments it received to the regions.

Had schools and regions been given clear financial incentives to have more sitters and passers at the outset of the pilot, their behaviour could have been influenced in various ways (potential consequences #4, 5, 6 and 7). In fact, however, the RBA incentives were more dilute at the school level than originally intended, as explained in previous chapters. As also noted earlier, the payments that were passed to schools were not provided in time to influence actions that could affect year 2 results and generally not in time to influence year 3 results either.

No data were available on grade repetition in EC 2006 (potential consequence #4).

Neither the verification team's report on the year 3 data nor the evaluation team's interviews at all levels of the education system finds evidence that schools were more likely to assist their pupils to cheat (potential consequence #5)<sup>73</sup>. Similarly, some schools may yet begin to focus on assisting pupils at risk of missing a pass by one or two grades on individual subjects at the expense of others where they are rewarded for passes (potential consequence #6), but analysis of the distribution of sitters' precise GPAs does not suggest that this has happened. Although there may be exceptions, interviews with teachers did not identify increasing temptations to move to schools rewarded under the scheme at the expense of needy schools (potential consequence #7). In fact, it is not obvious that this would be a major factor in teachers seeking to change schools or that they would be successful in being assigned to a better-performing school.

The potential for children from poor households and pastoralist communities to fall further behind others (potential consequence #8) could be realised if regions, woredas, and schools focused on 'easier wins' in seeking faster and easier progress towards more EGSECE sitters and passers. It is not clear, however, that the resources available in the pilot have been sufficient to trigger such responses. In principle, if the pilot continues, the problem could reduce outreach efforts and alter school admissions and levels of support in schools. Such changes would be difficult to detect because detailed school, community, and household data would be needed.

In conclusion, the evaluation team did not find evidence that any of the potential unintended consequences identified previously have occurred. Nor have other unintended consequences been observed. These findings should not surprise as reward payments were made only recently and financial incentives have been diffused.

---

<sup>73</sup> DFID's 2014 Annual Review of the RBA pilot concluded that 'The scope of verification does not, for example, detect teachers helping students during the exams...there are strong anti-cheating measures already in place, which countermand this incentive – including invigilation of exams by teachers from different schools and entire invalidation of school results by the National Examinations Agency, if irregularities are confirmed. Funds have not yet triggered strong incentives at school level, as RBA finance is only now starting to reach schools. There is, as a result, no evidence of such RBA-encouraged cheating happening'.

## 9 Summary, Key Lessons, and Recommendations

### 9.1 Summary

Before considering the team's recommendations it is first useful to summarise the findings associated with each of the seven issues identified in DFID's TOR for the evaluation. Doing so thus provides an entrée and context for the recommendations that follow.

**Results/Value added: To what extent did the RBA pilot increase educational results (compared with other/traditional methods including Ethiopia's General Education Quality Improvement Programme and Promoting Basic Services)? Are these changes attributable to RBA?**

There is no evidence in the estimates presented here that the RBA pilot improved educational performance for either boys or girls, in either the emerging regions or the non-emerging regions, in any year of the pilot.

The reason for lack of any detectable effect (beyond the measurement and estimation challenges) is not difficult to discern. First, the RBA scheme had not been communicated to the regions in time to affect students' performance appreciably during the pilot's three years. When the evaluation team visited regions in the pilot's first two years few of their education officials were aware of the pilot. When the verification team surveyed schools following the EC 2006 EGSECE, only about half the head teachers they interviewed knew about the pilot<sup>74</sup>.

Second, given the nature of the grading system, there is no reason to expect RBA to have any effect on the number of students passing the examination other than through its effect on the number sitting for it. For this reason, the evaluation team recommends that this indicator not be used in the future as a criterion for RBA payments.

There is, however, some reason to believe that RBA could affect the number of students sitting for the examination if the pilot were to be continued. RBA funds have begun to flow to regions and schools and, as a result, many regional and local staff are aware of the RBA incentive. Improvements to the educational system are being planned and some are being implemented. These improvements may well affect the number of students sitting for the EGSECE in future years.

**Equity: Who benefited from these improved results? Have disparities (boys versus girls, emerging regions versus others) declined?**

In addition to considering the overall effect of the pilot on the numbers of sitters and passers, the evaluation also assessed whether the pilot successfully reduced inequities related to gender or those between emerging and non-emerging regions. DFID had anticipated extra progress for girls compared to boys, and this was reflected in the higher reward payments offered for girls than for boys. Most important, there is no evidence that the RBA pilot accelerated progress toward gender equity. The increases that occurred were essentially a continuation of pre-existing trends rather than a change that can be attributed to the RBA pilot. During the pilot some of the gender inequities were reduced, but clearly not eliminated. This finding is not surprising. Addressing and eliminating the multiple and long-standing causes of the inequity within only three years is clearly an implausible goal.

---

<sup>74</sup> Coffey International Development, 'Phase 4 Verification Report, DFID Ethiopia Independent Verification of Key Government of Ethiopia (GOE) Educational Data for a Pilot of Results-Based Aid', 2015.



Similar findings apply to regional inequities. Slightly more than 10 per cent of all 15 and 16 year olds in Ethiopia are estimated to live in the four emerging regions. Accordingly, regional parity would be achieved when a similar percentage of all sitters and passers come from these four regions. The evaluation found no evidence of statistically significant deviations from baseline trends in the numbers of students sitting or passing the examination in either the emerging or non-emerging regions. While there were some gains in the emerging regions during the pilot's three years, the gains were not due to the pilot. Moreover, the gender gap appears to have closed more slowly in emerging regions than elsewhere and female EGSECE sitters continue to be considerably less likely than males to achieve high grades.

**System effects: Effects on resource allocation, education information systems, accountability, financial management, corruption)**

Advocates of RBA believe that reliance on country-based systems should not only strengthen these systems but also create incentives to improve them. In addition, according to DFID, reliance on RBA would similarly strengthen the government's accountability to its citizens and improve public financial management in the education sector. Given the findings noted above, one would not expect to see major or perhaps even discernible effects on these systems. While there may have been change at the margins, the evaluation was not able to identify any meaningful, pilot-related changes in: (a) the GoE's allocation of resources to lower secondary education; (b) the administration of the EGSECE; (c) the MoE's education management information system; (d) financial management; (e) accountability; or, (f) levels of corruption.

**Understanding why the approach works (or does not work): What factors and processes have been responsible for the results?**

Although the pilot did not accomplish as much as had been expected, there is evidence of RBA's potential to improve the quality of education in the future. RBA was being used to fund some regional and school-level changes during the pilot's third year and more RBA-funded changes were planned for the year after<sup>75</sup>. Furthermore, the strategic thinking of REBs taking the regional approach and schools in those regions taking the school approach indicates their willingness and ability to use discretionary funding to target specific needs.

Perhaps the RBA pilot assumed too much could be achieved in just three years. By the end of the pilot, there was evidence of its potential to increase the quality of secondary education. Moreover, there is evidence of RBA being seen as forward looking inasmuch as changes, within the limits of the regional RBA rewards, were being made at the regional and school levels with the expectation that they would increase the number of EGSECE sitters and passers and so generate further funding.

**Impact on aid relationships: Has the nature of the DFID/GoE dialogue improved?**

DFID hypothesised that the RBA pilot, with its 'hands-off' approach to development funding, would both change and improve the department's relationship with the GoE. DFID's aspiration that RBA might improve the dialogue between DFID and the MoE has not yet been realised. Although the MoE initially expressed enthusiasm, this positive sentiment was offset by concerns about the nature of the pilot itself. These concerns include a perception of high transaction costs, the relatively small amounts of money available through the pilot, and the absence of resources to pre-fund activities that might increase the numbers of sitters and passers. Interviews with DFID officials likewise reveal recognition that there has not been the anticipated behaviour change on the part of the MoE about RBA.

---

<sup>75</sup> DFID extended the RBA pilot for one year (i.e., to include the results of the EGSECE in EC 2007) but not the evaluation, which focuses only on the pilot's first three years.

More generally, other development partners active in the education sector have adopted a low-key yet polite response, essentially saying 'Let's see how this works out in practice, then we will consider our position'.

**Value for money: Do the benefits of the programme outweigh the costs? Are RBA incentives set at the optimal level?**

In the absence of any statistically significant effects of the pilot on the number of passers and sitters it is not possible to conclude that the RBA instrument has offered VfM to date. Nonetheless, the analysis does suggest that RBA as an instrument is a low-cost, relatively low-risk approach because its costs are low if it does not work (it has modest transactions costs if there are no reward payments) but with potentially high rewards albeit with higher costs if it does work.

For purposes of simplicity and transparency DFID chose to use a rolling baseline rather than an estimated counterfactual to calculate reward payments. As such, DFID risked having to pay for the counterfactual (results that would likely have been achieved in the absence of the RBA pilot). The team estimated the payments that DFID would have made using different baselines. The team did this using the counterfactual estimated as part of the design process (but not ultimately used in the final design) and a range of alternative baselines the team identified. In each of these alternative baselines DFID would have provided less money than the £15.6 it actually did, with the amount ranging from £2 million to £11 million depending on the scenario. The reward payments would also have had a different time frame.

Notwithstanding this finding, it is important to emphasise that there are no indications that the reward payments have been wasted or misused. While the payments have not yet contributed to the desired outcomes, Ethiopia's education system has received over £15 million that it would not have received in the absence of the pilot.

**Unintended consequences: To what extent has the programme resulted in unintended consequences? Have these been positive or negative? Were they identified early and remedial actions taken as necessary?**

New approaches to aid bring with them the possibility of unintended or unforeseen consequences. Such is the case with the RBA pilot. The evaluation team identified several possible unintended consequences, including: (a) more cheating to increase the number of sitters and passers; (b) pass rates increasing due to changes in the administration of the EGSECE; (c) schools focusing attention on students near the EGSECE pass/fail threshold at the expense of others; and, (d) increasing pressures for corruption. Despite the multiple opportunities for unintended consequences, no such occurrences were identified. Given the limited response to the RBA pilot by the end of year 3 and the lack of compelling evidence for the impact on the pilot's *intended* consequences, it would be surprising to find evidence of *unintended* consequences in the same period.

## **9.2 Another Perspective on the RBA's Results**

The summary has provided a sense of what the RBA pilot did and did not accomplish, but it is also useful to place the pilot in a broader perspective and to assess the results against widely used standards. The United Kingdom's Independent Commission on Aid Impact has provided such standards. In 2011 the commission initiated an assessment of DFID-sponsored education programmes in three East African

countries, including Ethiopia<sup>76</sup>. The ToR for the assessment included a series of questions focusing on four themes: objectives, delivery, impact, and learning. To promote comparability with that assessment Appendix 8 addresses each of these questions in the context of the RBA pilot.

### 9.3 Two Key Lessons

This report has provided many lessons and they have been discussed in their respective chapters. There are, however, at least two key lessons that merit further consideration.

The first lesson is that while RBA may be conceptually appealing, its transformation into practical policy and implementation is not an easy task. A natural and understandable resistance to change exists in all bureaucracies and particularly among government ministries and departments. They typically face a plethora of rules and regulations that limit or proscribe their administrative flexibility. Major change is not likely to happen quickly, especially when the change involves new, novel, and experimental approaches to the provision of public services and when the perceived benefits of the change are, as the *Business Case* observed, 'very difficult to predict'.

The potential appeal of the change is likewise diminished when the resources associated with the change are minimal. Consider the first year of the RBA pilot. DFID offered as much as £10 million for results achieved in EC 2004, but the actual reward payment was less than 10 per cent of this amount and only a microscopic fraction of the GoE's overall education budget. Advocating the merits of a programme such as the RBA pilot would be difficult under any circumstances, but especially when the anticipated beneficiaries, Ethiopia's secondary schools and their students, are unsure whether their efforts will produce the desired results. Under ideal circumstances all Ethiopian schools would have electricity, running water, and toilets for girls and boys; many do not. Providing these resources will improve the quality of student's educational experience but not necessarily the number of EGSECE sitters and passers.

The RBA pilot was further based on the assumption that many of the barriers to increased numbers of sitters and passers are amenable to actions *within* schools and can be addressed quickly, such as within the three years of the pilot. In many instances, in contrast, the barriers reflect long-standing cultural norms and are outside schools' control or influence. In Afar, for example, the gross enrolment ratio (GER) in lower secondary schools in EC 2006 was only 8.5 per cent (and was even lower for girls). Among Ethiopia's 11 regions only four had overall GERs above 50 per cent and only two had GERs above 50 per cent for girls in that year<sup>77</sup>. If schools cannot enrol girls and boys in their catchment areas, then there is little reason to expect meaningful increases in the number of sitters and passers before problems with enrolment and retention are resolved.

Moreover, the causes of this and other problems with the educational system are long standing. It is unreasonable to assume that RBA can remedy or mitigate them in only three years. Perhaps for this reason the Center for Global Development considers that five years should be the minimum length for an RBA initiative<sup>78</sup>.

The second key lesson involves DFID's expectations for evaluation. A theory of change posits a series of *assumptions*, all of which merit examination. As hundreds of examples demonstrate, what appears

<sup>76</sup> ICAI, 'DFID's Education Programmes in Three East African Countries', ICAI, Report 10, May 2012. Available at <http://www.oecd.org/countries/rwanda/50360183.pdf>

<sup>77</sup> MoE, *Education Statistics Annual Abstract, EC 2006*. The GER is the number of students enrolled divided by the school age population.

<sup>78</sup> Nancy Birdsall and William D. Savedoff. 2010. *Cash on Delivery: A New Approach to Foreign Aid* (Washington, DC: Center for Global Development). Available at [http://www.cgdev.org/sites/default/files/1423949\\_file\\_CODAid\\_SECOND\\_web.pdf](http://www.cgdev.org/sites/default/files/1423949_file_CODAid_SECOND_web.pdf).

attractive in theory does not always assure achievement of desired or anticipated results. This important distinction both underlies and justifies the need for rigorous evaluation that employs a valid counterfactual. Without this counterfactual it is not possible to compare the effects of the intervention, in this case the RBA pilot, with the absence of the intervention or to conclude with confidence that the intervention and its investment are responsible for any of the changes observed. Such is the case with the RBA pilot. The GoE was understandably interested in including all of its regions in the pilot, but the choices that DFID and the MoE made in designing and implementing the pilot precluded a well-designed impact evaluation with a valid counterfactual – despite DFID’s recognition of its desirability.

In fact, a case can be made that much of the evaluation *preceded* the project’s implementation and perhaps was conducted prematurely. This evaluation discusses ‘results’ from EC 2004 and EC 2005. For both years, however, the key actors and assumed implementers, schools and REBs, were largely unaware that the pilot existed. The initial rewards flowed to them toward the end of EC 2006 and only after all three cycles of the EGSECE had already been completed.

This report has provided what the evaluation team considers to be considerable information of value, but several important issues remain unaddressed. For example, a fundamental assumption of RBA is that the prospect of reward payments for outcomes anticipated to be achieved in the future will provide sufficient incentive for governments to marshal the resources necessary to fund the initiatives that lead to the desired outcomes<sup>79</sup>. The validity of this assumption should be confirmed. If this assumption is invalid, in contrast, then it undermines the theory of change as well as the appeal and likely effectiveness of RBA. Would Ethiopia’s REBs and secondary schools have pre-funded relevant interventions if they had known at the beginning of the pilot that success would subsequently provide them with financial rewards? The evaluation team was not able to answer this question because of the pilot’s delayed implementation.

Despite using what are arguably the best possible alternatives to an impact evaluation, the present evaluation may also not fulfil the needs or the reasonable expectations of the report’s intended audience and key stakeholders. To illustrate, a review of the draft evaluation report commented that the (a) pilot’s ‘emerging results, in terms of responses to incentives are promising, and if the pilot had been given more time the theory of change for the RBA may well have developed strength’ and (b) pilot ‘likely will lead to increased education outcomes, and likely represent extremely good VFM’. The evaluation team hopes these expectations are eventually achieved, but the challenges associated with the evaluation as well as its limitations require that these statements be considered as aspirational rather than evidence based. Given (a) the resources devoted to the RBA pilots in Ethiopia and Rwanda; (b) DFID’s acknowledgment in the *Business Case* that RBA is ‘untested’ and that a ‘robust evidence base’ is absent; (c) its intention to increase its reliance on RBA in other countries and sectors; and, (d) the department’s desire that other donors and stakeholders buy in to the RBA approach, opportunities for well-designed and methodologically sound evaluation should not be neglected.

## 9.4 Recommendations

1. If the pilot were to be continued to year 5, DFID and GoE should increase the premiums for girls and the emerging regions (see Chapter 3 and Table 3.1) based on the analysis of the challenges they continue to face and the use made of the RBA rewards to date. Similarly, DFID and the GoE should decide whether rewards should be distributed according to results achieved, as was the original intention, or with the purpose of regional equity, and so only partially related to performance and outcomes achieved.

<sup>79</sup> DFID, *Business Case* and Nancy Birdsall and William D. Savedoff. 2010. *Cash on Delivery: A New Approach to Foreign Aid*.

2. Recognise that, in switching to an RBA approach, expectations of speedy changes in donor and recipient behaviour must be set, on both sides, against the background of many years of working in completely different and often incompatible ways. Behavioural change takes time in any setting. Do not underestimate the work and time) involved in explaining and testing the understanding of new approaches.
3. Recognise that vigorous local ownership and engagement with a new approach such as RBA are a prerequisite and ensure that awareness, buy in, and ownership exist among all levels of stakeholders before implementation begins.
4. When working in a new context, recognise that transaction costs for verification, communication, monitoring, and evaluation are likely to be high. RBA does not create capacity to administer a reward-based system, but such capacity is a prerequisite for success.
5. Payments should be based on increases above the trend over some previous period (perhaps five years) rather than the change from the previous year. The numbers of sitters and passers have been on an upward trend in Ethiopia. The result is that under the current system DFID is paying for some increases that would have occurred even in the absence of RBA. This does not provide value for money. Under the current system, a region that shows an increase in sitters or passers in one year, a decrease in the next, and an equal increase in the third year will receive payments in years 1 and 3 – in effect, being paid twice for the same increase in sitters or passers. This feature of the payment formula was largely responsible for the large payment for results in EC 2005, as regions rebounded from the decreases in sitters in EC 2004. Basing payments on increases above trend would eliminate this anomaly. The trend would need to be constructed carefully, considering other factors such as enrolments in feeder grades as well as the raw numbers of sitters and passers. There would also be a periodic need to review whether the trend continued to apply or needed to be adjusted.
6. Financial and capacity-building pump-priming should be considered to allow the initial progress to be made and rewarded. This is important if trends are used rather than adjusting baselines, as business as usual will not deliver an initial reward payment. In a resource-constrained environment, such as Ethiopia's education sector, few resources are likely to be available to 'prime the pump' and thus to fund or support meaningful changes that lead to increases in sitters and passers. This is especially true in the absence of a clear and definitive linkage between the considerable number of possible investments in education and the desired results.
7. Reward payments should not be based on the number of passers of norm-referenced examinations like the EGSECE. The EGSECE results are 'normalised' each year (i.e., rescaled to hold the mean constant from year to year) with the consequence that the examination is a poor measure of educational progress<sup>80</sup>. Year-to-year variation in the number of students who pass the EGSECE reflects a complex interaction of variation in the difficulty of the test and the way in which the NEAEA handles students with raw scores obtained from pure guessing on a wholly multiple-choice examination like the EGSECE. Reward payments for sitters could be increased to balance the cessation of reward payments for passers.
8. In relation to VfM, recognise that the context of RBA is new, with no right and wrong ways of thinking about VfM, only different interpretations. As the evidence base gets stronger, robust VfM

---

<sup>80</sup> The evaluation team is not declaring the EGSECE to be a poor examination for the purposes for which GoE employs it, namely decisions about who will be eligible to progress to upper secondary education, but the EGSECE cannot be used to measure educational progress from year to year.

methodologies will develop. Similarly, if DFID wishes to compare the cost effectiveness of its interventions it should ensure the use of common measures and methods across the interventions.

## 10 Bibliography

- Allsop, T., Ellison, R., Mundy, K., Orr, L., Pearson, M., Ridley, B., Vohra, J., and Candid Services. 2012. 'Evaluation of the Pilot Project of Results-Based Aid in the Education Sector in Ethiopia – Inception Report' (Cambridge Education: Cambridge).
- Allsop, T., Ellison, R., Mundy, K., Orr, L., Pearson, M., Ridley, B., Vohra, J., and Candid Services. 2013. 'Evaluation of the Pilot Project of Results-Based Aid in the Education Sector in Ethiopia – Baseline Report' (Cambridge Education: Cambridge).
- Allsop, T., Ellison, R., Mundy, K., Orr, L., Pearson, M., Ridley, B., Vohra, J., and Candid Services. 2014. 'Evaluation of the Pilot Project of Results-Based Aid in the Education Sector in Ethiopia – Year 1 and 2 Report' (Cambridge Education: Cambridge).
- Asgedom, A. and Ridley, B. 2015. 'Historical Narratives in Ethiopia' in Paul Smeyers, et al., ed., *International Handbook of Interpretation in Educational Research*.
- Birdsall, N. and W.D. Savedoff. 2010. *Cash on Delivery: A New Approach to Foreign Aid* (Washington, DC: Center for Global Development). Available at [http://www.cgdev.org/sites/default/files/1423949\\_file\\_CODAid\\_SECOND\\_web.pdf](http://www.cgdev.org/sites/default/files/1423949_file_CODAid_SECOND_web.pdf), accessed 14 August 2015.
- Birdsall, N. and Perakis, R. 2012. 'Cash on Delivery Aid: Implementation of a Pilot in Ethiopia'. Available at [http://www.cgdev.org/sites/default/files/archive/doc/Initiatives/Ethiopia\\_RBA\\_pilot\\_report.pdf](http://www.cgdev.org/sites/default/files/archive/doc/Initiatives/Ethiopia_RBA_pilot_report.pdf), accessed 5 August 2015.
- Center for Global Development. 2014. 'The Anatomy of Program-for-Results: An Approach to Results-Based Aid', Working Paper No. 374. Available at [http://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=2466657](http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2466657), accessed 5 August 2015.
- Coffey International Development. 2012. 'Baseline Report' (Reading: Coffey International Development).
- Coffey International Development. 2012. 'Inception Report' (Reading: Coffey International Development).
- Coffey International Development. 2013. 'Phase 2 Verification Report: Independent Verification of Key Government of Ethiopia (GoE) Educational Data for a Pilot of Results-Based Aid (RBA)' (Reading: Coffey International Development).
- Coffey International Development. 2014. 'Phase 3 Verification Report: Independent Verification of Key Government of Ethiopia (GoE) Educational Data for a Pilot of Results-Based Aid (RBA)' (Reading: Coffey International Development).
- Coffey International Development. 2015. 'Phase 4 Verification Report: Independent Verification of Key Government of Ethiopia (GoE) Educational Data for a Pilot of Results-Based Aid (RBA)' (Reading: Coffey International Development).
- Dee, T.S. and Jacob, B.A. 2011. 'The Impact of No Child Left Behind on Student Achievement', *Journal of Policy Analysis and Management*, 30, 3, pp. 418-446. Available at <http://www.nber.org/papers/w15531>, accessed 5 August 2015.



DFID. No date. 'Guidance on Aid Instruments: A DFID Practice Paper'. Available at <http://www.mtnforum.org/sites/default/files/publication/files/1873.pdf>, accessed 5 August 2015.

DFID. 2011. 'DFID's Approach to Value for Money'. Available at [https://www.gov.uk/government/uploads/system/uploads/attachment\\_data/file/67479/DFID-approach-value-money.pdf](https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/67479/DFID-approach-value-money.pdf), accessed 5 August 2015.

DFID. 2011. 'Pilot Project of Results-Based Aid (RBA) in the Education Sector in Ethiopia: Business Case'. Available at [iati.dfid.gov.uk/iati\\_documents/3716785.docx](http://iati.dfid.gov.uk/iati_documents/3716785.docx), accessed 5 August 2015.

DFID. 2011. 'Terms of References (TOR) Evaluation of Pilot Project of Results-Based Aid (RBA) in the Education Sector – Ethiopia'.

DFID. 2013. 'Annual Review: Pilot Project of Results Based Aid (RBA) in the Education Sector in Ethiopia'. Available at [iati.dfid.gov.uk/iati\\_documents/4341561.docx](http://iati.dfid.gov.uk/iati_documents/4341561.docx), accessed 5 August 2015.

DFID. 2014. 'Annual Review: Pilot Project of Results-Based Aid (RBA) in the Education Sector in Ethiopia'. Available at [iati.dfid.gov.uk/iati\\_documents/4839826.docx](http://iati.dfid.gov.uk/iati_documents/4839826.docx), accessed 5 August 2015.

DFID/Ethiopia. 2015. 'Terms of Reference (TOR) for Contracting an individual consultant for Independent Verification of Key Government of Ethiopia (GoE) Educational Data of 2014/15 for a Pilot of Results-Based Aid (RBA) – Ethiopia'.

GoE, Ministry of Education. 2010-2015. *Education Statistics Annual Abstracts*. Available at <http://www.moe.gov.et/English/Information/Pages/AnnualAbstract.aspx>, accessed 5 August 2015.

GoE, Ministry of Education. 2012. 'Results Based Aid (RBA) Pilot Guidelines: Improving access, quality and equity in general secondary education', EC 2004.

GoE, Ministry of Federal Affairs. 2007. 'Emerging Regions Development Programme'. Available at [http://www.unctf.org/sites/default/files/Documents/erdp\\_54573\\_prodoc\\_0.pdf](http://www.unctf.org/sites/default/files/Documents/erdp_54573_prodoc_0.pdf), accessed 5 August 2015.

GoE, National Educational Assessment and Examinations Agency. 2010. 'Ethiopia First National Learning Assessment of Grades 10 and 12 Students'.

Gultie, T., Hailu, D., and Workineh, Y. 2014. 'Age of Menarche and Knowledge about Menstrual Hygiene Management among Adolescent School Girls in Amhara Province, Ethiopia: Implication to Health Care Workers & School Teachers'. Available at <http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0108644>, accessed 5 August 2015

HIFAB International. 2013. 'Comprehensive Evaluation of the General Education Quality Improvement Programme (GEQIP) – Draft Exit Survey Report'.

HIFAB International. 2013. 'Comprehensive Evaluation of the General Education Quality Improvement Programme (GEQIP) – Final Exit Survey'.

Independent Commission on Aid Impact. 2012. 'DFID's Education Programmes in Three East African Countries' Available at <http://www.oecd.org/countries/rwanda/50360183.pdf>, accessed 5 August 2015.

ITAD. 2015. 'Value for Money Assessment of the Protection of Basic Services Programme'.

Kenny, C. and Savedoff, W. 2013. 'Can Results-based Payments Reduce Corruption?' Center for Global Development, Working Paper 345. Available at <http://www.cgdev.org/publication/can-results-based-payments-reduce-corruption-working-paper-345>, accessed 5 August 2015.

Organisation for Economic Co-operation and Development. 2002. *Glossary of Key Terms in Evaluation and Results Based Management*. (Paris: OECD). Available at <http://www.oecd.org/dac/evaluation/glossaryofkeytermsinevaluationandresultsbasedmanagement.htm>, accessed 5 August 2015.

Schochet, P.Z. 2008. 'Technical Methods Report: Guidelines for Multiple Testing in Impact Evaluations (NCEE 2008-4018) (Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education). Available at <http://ies.ed.gov/ncee/pdf/20084018.pdf>, accessed 5 August 2015.

Tegegne, T. and Sisay M. 2014. 'Menstrual hygiene management and school absenteeism among female adolescent students in Northeast Ethiopia'. Available at [www.biomedcentral.com/1471-2458/14/1118](http://www.biomedcentral.com/1471-2458/14/1118), accessed 5 August 2015.

Upper Quartile. 2014. 'Evaluation of Results-Based Aid in Rwandan Education: Econometric Report 2014 (draft)'.

Vaughan, S. 2015. 'Going against the Grain? Lessons from a DFID Results-Based Aid pilot in Ethiopia', Policy brief completed for DFID/Ethiopia.

Wong, M., Cook, T.D., and Steiner, P.M. 2009. 'No Child Left Behind: An Interim Evaluation of its Effects on Learning Using Two Interrupted Time Series Each With Its Own Non-Equivalent Comparison Series'. Available at <http://www.ipr.northwestern.edu/publications/docs/workingpapers/2009/IPR-WP-09-11.pdf>, accessed 5 August 2015.