upperquartile



Final Evaluation Report

Evaluation of Results Based Aid in Rwandan Education

December 2015

Submitted by

Upper Quartile

in association with

Institute of Policy Analysis and Research - Rwanda

Institute of Policy Analysis and Research - Rwanda

Contents

List of acronyms iii
Glossary of terms v
Executive Summary vi
1. Introduction and background 1
1.1. Evaluation purpose and scope of work 1
1.1.1. Purpose of the evaluation 1
1.1.2. Scope of the evaluation 1
1.1.3. Evaluation questions 1
1.1.4. Evaluation timing 2
1.1.5. Evaluation audiences 2
1.1.6. Deviations and additions to the original terms of reference 2
1.2. Understanding the context 3
1.2.1. Conceptualising payment by results: the theoretical framework 3
1.2.2. The policy drivers 4
1.2.3. Payment by Results in Rwanda 5
1.3. The evaluation findings to date 6
2. Methodology 7
2.1. Methodological approach 7
2.2. The evaluation framework in year three 7
2.3. Methods 7
2.3.1. Quantitative methods 8
2.3.2. Qualitative methods 10
2.3.3. Analysis and triangulation 10
2.4. Limitations of the evaluation 12
2.4.1. Limitations of the econometric evidence 12
2.4.2. Limitations of the Value for Money evidence 12
2.4.3. Limitations of the qualitative evidence 12
2.5. Research ethics 12
3. The evaluation findings 13
3.1. Completion 13
3.1.1. RBA Payments 13
3.1.2. A significant change? 17
3.1.3. The explanatory factors for change 19
3.2. English language 24
3.2.1. RBA Payments 24
3.2.2. A *real* change? 25
3.2.3. The explanatory factors for change 26
3.3. RBA as a factor contributing to change 28

3.3.1. Completion 29
3.3.2. English language proficiency 32
3.3.3. Unanticipated or unintended effects 37
3.4. Value for Money 39
3.4.1. What is Value for Money in relation to RBA? 39
3.4.2. An approach to assessing VfM 39
3.4.3. RBA in Rwandan Education – Value for Money? 42
3.4.4. RBA in Rwandan Education – Do the assumptions stack up? 44
4. Conclusions and recommendations 45
5. References 49
APPENDICES 51
Appendix 1 – Terms of Reference for Evaluation of Project of Results Based Aid in the Education Sector – Rwanda 50
Appendix 2 – Terms of reference for technical assistance to explore options for assessing value for money of the Results Based Aid pilot in the Education Sector 66
Appendix 3 – 2015 Research Approach Paper 70
Appendix 4 – List of those consulted 82
Appendix 5 – Research tool [National level KIs] 83
Appendix 6 – Research tool [District/ Sector-level officials] 86
Appendix 7 – Focus Group Guide [Principals] 88
Appendix 8 – Report of the econometric modelling 89
Appendix 9 – Detailed report of the Value for Money Assessment 104

Report prepared by Upper Quartile:

Paul Clist (Quantitative Research Lead), **Brendan Whitty** (Qualitative Research Lead), **Joseph Holden** (Value for Money Lead), **Pamela Abbot** and **IPAR-Rwanda** (District and School-Level Research team), **Kathleen Latimer** (UQ Evaluator and Project Manager), **Kyla Reid** (UQ Evaluator) and **Chris Boyd** (UQ Project Director).

List of acronyms

12 YBE	Twelve year basic education
9 YBE	Nine year basic education
BCR	Benefit Cost Ratio
CEFR	Common European Framework for Reference
CGD	Center for Global Development
COD	Cash on Delivery aid
DEO	District Education Officer
DFID	Department for International Development
DG	Director General
DP	Development Partner
EC	European Commission
EDC	Education Development Centre
EMI	English as the Medium of Instruction
EMIS	Education Monitoring Information System (data)
ESSP	Education Sector Strategic Plan
FGD	Focus Group Discussion
GBP	Great British Pound (£)
GOR	Government of Rwanda
HEART	Health, Education and Advice Resource Team
HMG	Her Majesty's Government
ICT	Information and communications technology
IEE	International Education Exchange, an NGO
IPAR – Rwanda	Institute of Policy Analysis and Research – Rwanda
JRES	Joint Review of the Education Sector
KI	Key Informant
KII	Key Informant Interview
L3	[Rwandan] Literacy, Language, and Learning Initiative
M&E	Monitoring and Evaluation
MINALOC	Ministry of Local Government
MINECOFIN	Ministry of Finance and Economic Planning
MINEDUC	Ministry of Education

MOU	Memorandum of Understanding
NAO	[United Kingdom] National Audit Office
NGO	Non-Governmental Organisation
(N)PV	(Net) Present Value
P1, P2 etc.	Primary 1, Primary 2 etc
PbR	Payment By Results
PEA	Political Economy Analysis
RBA	Results Based Aid
RBF	Results Based Finance
REB	Rwanda Education Board
RESP	Rwanda Education Sector Plan
RPI	Retail Price Index
S1, S2 etc	Secondary 1, Secondary 2 etc
SBM	School Based Mentor
SBMP	School Based Mentor Programme
SBS	Sector Budget Support
SEO	Sector Education Officer
SMT	[Government of Rwanda, Education] Senior Management Team
SWAp	Sector Wide Approach
TOC	Theory of Change
TOR	Terms of Reference
TTC	Teacher Training Course
TVET	Technical and Vocational Education and Training
UPE	Universal Primary Education
UQ	Upper Quartile
UKAID	UKAID and the associated logo is used widely in the countries where DFID works to refer to and acknowledge development programmes funded by the UK taxpayer
USAID	United States Agency for International Development
VfM	Value for Money
VSO	Voluntary Service Overseas

Glossary of terms

The aim in this report is to strike a balance between providing sufficient detail so as to allow an assessment of the quality and breadth of the technical work undertaken, and providing sufficient clarity so as to allow a non-specialist reader to understand the key messages in the report. In order to aid the non-technical reader a glossary of key terms is provided here. Throughout the text, SMALL CAPS are used to denote a term that appears below.

- **COEFFICIENT–** The estimated COEFFICIENT describes the strength of the effect that a one unit increase in the INDEPENDENT VARIABLE has on the DEPENDENT VARIABLE.
- **DEPENDENT VARIABLE –** In crude terms, the thing we are trying to explain.
- **DUMMY–** A VARIABLE which takes the value 1 when a condition is met, and 0 otherwise. For example, a year dummy for 2012 takes the value 1 when the year is 2012, and 0 otherwise. In this example, the coefficient would measure the effect of it being 2012 relative to the base year.
- **FIXED EFFECTS –** An effect which is constant across the individuals in the group being examined, and can therefore be controlled for in our model.
- **HOMOSCEDASTICITY –** This describes a situation in which the error term (that is, the "noise" or random disturbance in the relationship between dependent variable and the independent variables) are consistent across the model. For example, in one model we would assume that errors are not much larger in districts with higher completion rates.
- **INDEPENDENT VARIABLE(S) –** In crude terms, the thing(s) we are using to explain the DEPENDENT variable.
- **NET PRESENT VALUE (NPV) –** The total current value of future activities, minus any related costs
- **OBSERVATION –** One data point. If we have data on the population of each Rwandan district for one year, we have 30 OBSERVATIONS (as there are 30 districts). If we have it for two years, we have 60 observations.
- **OUT-OF-SAMPLE PREDICTION –** This helps us test the accuracy of our model. We first run the model on a subset of data, deliberately excluding some OBSERVATIONS (specific years or districts). We then use those COEFFICIENTS to 'predict' the excluded OBSERVATIONS. We can then compare the prediction with the known outcome.
- **PRESENT VALUE –** The total current value of future activities. For example, £200 in one year's time may have a present value of £150, as a person would be willing to trade off some of the amount in order to receive it earlier.
- **SIGNIFICANT –** The measure of how likely it is to see an effect purely through chance. To be significant at the 1% level means that once in 100 times you would see the effect and it would purely be due to chance. The other typical SIGNIFICANCE levels are 5% and 10%. The smaller the level of SIGNIFICANCE, the more confidence the evaluator can have in the evidence.
- **SPECIFICATION–** The list of INDEPENDENT VARIABLES included in a specific model and the type of estimation technique used.
- **STANDARD ERRORS –** A measure of the statistical accuracy of an estimate.
- **SUR MODEL –** Seemingly unrelated regressions (SUR) models are systems of regression equations, each of which has its own dependent variable and potentially different sets of exogenous explanatory variables. Each is a valid linear regression on its own, but their error terms are assumed to be correlated. This approach is useful in that it gives more efficient estimates by combining information on different equations.
- **VARIABLE –** An indicator or measurement, such as population or teacher numbers.

Executive Summary

Introduction

Upper Quartile was commissioned to undertake a three-year mixed-method process and impact evaluation of the Results Based Aid (RBA) pilot in Rwandan education, considering if, how and in what circumstances the pilot contributed to the results envisaged in the Memorandum of Understanding (MoU) between the Department for International Development (DFID) and the Government of Rwanda (GoR). This final evaluation report builds on the year one and two reports (Upper Quartile, 2014, 2015).

Purpose of the evaluation

The purpose of the evaluation is to determine whether the additional incentive of RBA impacted completion at key stages of education[1] and additional teachers becoming competent in the use of English as the medium of instruction.[2]

The evaluation also considers the response of the recipient (GoR) and other key actors to RBA and seeks to establish the various processes that led to observed results.

This final report presents findings in relation to 2014 completion data and data on English language proficiency, as well as findings from across the three year pilot.

Methodology

The methodological approach to the evaluation is 'realist'; exploring questions about what works, for whom, in what circumstances and why.

The **impact evaluation** component is premised on the findings of an econometric model which explores trends in, and the factors affecting, completion with the aim of identifying any RBA-related effect.

The **process evaluation** in each year involved desk-based review of policy and literature, combined with qualitative primary research at national, district and school-based level to explore the response to RBA and the drivers of change in relation to completion and English language proficiency. In year three, on the basis of the combined evaluation evidence base, [3] the qualitative research set out specifically to explore three research avenues which may indicate an effect of RBA – increased completion at S3 in 2013, increased completion at S6 in 2014 and improvements in English language proficiency of teachers as recorded in the endline assessment of language capability in 2014.

The evaluation also included in-depth modelling of the potential **value for money (VfM)** of RBA. This research will contribute to the theoretical debate around payment by results (PbR) mechanisms.

The evaluation seeks to address seven macro-evaluation questions posed in year one:

1) **What has been achieved?**
2) **Has the RBA approach contributed to impact in relation to the envisaged results?**
3) **What factors have impacted on the achievement of the RBA results?**
4) **How is the RBA approach perceived in Rwandan education?**
5) **How did government respond to RBA?**
6) **Has value for money been achieved?**
7) **What lessons have been learned to inform RBA in Rwanda and elsewhere?**

The mixed-method evaluation approach is summarised below in Table E1.

Table E1: The evaluation approach

Method	Approach
Econometric modelling	Modelling drawing on national level secondary data sources to identify any effect of RBA over and above what may have been expected (in terms of completion) in the absence of RBA.
VfM analysis	Considers the cost-effectiveness of RBA relative to not providing RBA. Using national level secondary data sources and standard practice for assessing VfM, two different counterfactuals were constructed.
Qualitative research	Semi-structured interviews were conducted with national level strategic stakeholders, district and sector officials and school Principals (in three districts) to explore specific research avenues identified for year three and the response of GOR to the RBA modality.

[1] 'Completion' is defined as additional learners sitting key stage examinations in the sixth grade of primary school (P6), the third and sixth grades of secondary school (S3 and S6).

[2] 'Competency' has been defined as additional teachers reaching level B1 proficiency in the Common European Framework for Reference (CEFR) scale. A baseline sample survey was undertaken by the British Council in 2012 with a follow-up sample survey administered in November/ December 2014.

[3] Including the econometric and VfM modelling in year three.

Payments

Over the three years of the pilot, £5,440,590 GBP was disbursed from a possible maximum of £9 million GBP; approximately 60% of available funds have been disbursed as RBA (although it is noted that the £9 million GBP ceiling figure was in not intended as a target).

Table E2: The exam results and payments

Pilot year	Exam sitters			English proficiency: % ≥ B-level
	P6	S3	S6	
2011	154,954	77,420	46,558	-
2012	166,153	80,590	50,302	3.2%
2013	163,094	93,732	53,791	-
2014	157,123	86,091	69,283	49%
Change 2011-14	+1%	+11%	+49%	-
Total payment	£775,940	£1,931,850	£1,554,700	£1,178,100

Completion findings

During the pilot period, there has been an increase at the higher grades in the numbers of students sitting exams, with virtually flat trends at P6.

Looking over the results from the pilot, there are two key completion "bumps" which require explanation. This includes SIGNIFICANT above trend completion at the S3 level in 2013 and SIGNIFICANT above trend completion at the S6 level in 2014. With the exception of S6, in 2014 there were no other positive SIGNIFICANT annual changes observed.

2014 S6 completion

The SIGNIFICANT above trend completion at S6 in 2014 is the result of a cohort effect. Examining available enrolment data, it is clear that a large group of students enrolled in S2 in 2010. Four years later, in 2014, this group of students sat the S6 exam. The 2014 S6 cohort would have enrolled in S1 in 2009—the year 9-year Basic Education (9YBE) was introduced in Rwanda. This conclusion is supported by overwhelming consistency in interview responses. The 2014 S6 increase was linked by every interviewee, both at the national level and in the districts, to the introduction of fee-free schooling; particularly 9YBE in 2009 (and subsequently 12-year Basic Education (12YBE) in 2012.

Examination of the transition rates adds further support—the 2013-14 transition rates of students from S5 to S6 are comparable to other years, implying that the high number of exam sitters was heavily influenced by the number of students enrolled before the RBA MOU was agreed. Therefore, it was historical policy changes, not RBA, which caused the 2014 S6 increase (see section 3.1.3.1).

2013 S3 completion

The increased in completion at S3 in 2013 is more difficult to explain. In the year two report analysis suggested some of the increase was attributable to a cohort effect. The analysis in year three adds more data which allows the econometric modelling to control more for cohort effects. This has revised down the estimate of "additional" completers, but there remains a SIGNIFICANT and positive increase. Therefore, a cohort effect can only partly explain this jump.

The qualitative research in year three highlighted three explanatory factors: 1) lagged uptake of 9YBE, linked to the growing trust in the quality of the education system and the school construction programme; 2) community and school-led efforts to reduce drop-out; and 3) a directive from REB to the effect that senior secondary students without a certificate from S3 must come back and sit for the S3 exam.

Of these, RBA is not considered as a potential influencer in either factor one or three. The first was part of a long-term GOR programme that was initiated long before RBA. Regarding the latter, there is no evidence that the administrative re-emphasis of an existing rule was caused by RBA (and any change would have been 'gaming' rather than a beneficial effect).

Looking at the second explanatory factor, the pecuniary incentive of RBA weakly/moderately incentivised communication efforts. RBA may have acted as a contributing incentivizing factor to the communication efforts from the centre which sought to focus the attention at district and sector-level on preventing drop-out. This may have been based in part on a misunderstanding of the RBA payment mechanism. This drive may have worked to add a very slight incentive effect, impacting a small number of students (see section 3.1.3.1).

English language findings

In 2012 the British Council undertook a survey of teachers to ascertain baseline levels of

English proficiency in relation to the six levels of the Common European Framework for Reference (CEFR). Results of the 2012 baseline survey showed that the vast majority of teachers (96.8%) possessed only a basic level of English language proficiency. The 2014 follow-up survey showed a marked improvement, with 43.4% of teachers assessed at the intermediate level compared to 2.9% in 2012.

Unfortunately, while the results appear impressive, there are a large number of concerns regarding the data. The 2012 and 2014 results are from different exams, taken under different conditions. Each of the changes to the 2014 exam are expected, by the British Council, to have made the test a more accurate measure of teacher's English, but also to have positively affected the results. Because of this, the results of the two tests are not comparable; it is not possible to know how much of the change in test results is due to any true improvement in the English proficiency of teachers, and how much is due to the use of a different test under different conditions.

We can examine the ways RBA may have affected teachers' English language proficiency, even if we cannot state what those effects were. RBA exerted influence in several different ways. It had a strong effect to increase the urgency of GOR policy-making, which moderately reinforced the sector leaderships' efforts to stabilise the focus on English proficiency. However, it had a mixed contribution to sector leaderships' coordination of the international actors (depending on the actor), and very weakly reinforced their communication efforts at the district and school level. It also seems likely that the positive endline results have negatively affected policy, by creating a (possibly false) impression of success (see section 3.2.3).

Value for Money Findings

The VfM models show there is a significant benefit from the application of the RBA mechanism, and due to the low break-even number of additional sitters, even a small-scale impact on completion could be VfM. However, these findings should be treated with caution, as the quantitative and qualitative findings cast significant doubt on the models' assumptions; particularly the extent to which the increase in completion is truly *additional*. This means that the statistically significant additional sitters modelled by the evaluation (the basis of the VfM assessment) are unlikely wholly additional; the result instead of chance, wider policy implementation, cohort effects or perverse incentives.

Further, there was some evidence from MINEDUC that achieving less than £3 million per year in disbursements was experienced as a short-fall, partly due to misunderstanding about how RBA would function. This uncertainty effect could offset any incentive effect generated.

The benefit derived from the VfM model estimated to come from DFID support to education in Rwanda through the mechanism of RBA should therefore be treated with caution. However if a desirable incentive effect (e.g. on educational quality) could be created through a PbR aid mechanism and negative effects avoided, the model shows this could easily generate very significant returns (see section 3.4.3).

Conclusions

We examine the evaluation conclusions through the evaluation questions.

1. **What has been achieved?**

In the context of the RBA indicators the most impressive achievement has been improvement in the absolute number of exam sitters at both S3 and S6 levels, with respective increases of 11% and 49%. Performance at P6 has been virtually flat. With regards to the English proficiency of teachers, any apparent success is undermined by inconsistent data: a different test was taken in different conditions. These achievements saw £5.4 million disbursed as RBA.

2. **Has the RBA approach contributed to impact in relation to the envisaged results?**
3. **What factors have impacted on the achievement of the RBA results?**

The quantitative evidence is unanimous in finding that RBA had no consistent effect on completion results. Where completion performance was outside of the range of what would be expected without RBA, performance was as likely to be below trend as above it. There were two instances of above-trend performance: S3 exam sitting in 2013 and S6 exam sitting in 2014. The latter performance was unanimously linked to the historical policy changes of 9YBE and 12YBE. The picture surrounding S3 performance in 2013 is less clear. There are three main proximate causes of an S3 increase. It is not possible to comment using quantitative data on the extent to which RBA may have impacted results for the indicator of English language proficiency due to

lack of comparability of data. In the qualitative evidence, four influencing factors in the area of English language proficiency were identified: the School Based Mentoring Programme, improved teacher training, the general management communication effort to the teaching profession around the importance of English language proficiency, and activities of other international donors and NGOs in this policy area.

4. How is the RBA approach perceived in Rwandan education?

The RBA agreement was perceived within senior elements of the GoR hierarchy as an experiment. Although officials were aware that DFID would disburse on completion and English-proficiency related targets, the precise mode of payment was not clear within key elements of the main implementing body, REB, until the last payment. The realisation that DFID disbursals under RBA were not earmarked for English language proficiency or completion but were part of the negotiation with MINEDUC, was disappointing for REB officials. More broadly, RBA is not understood well outside a small group of senior officials.

5. How did government respond to RBA?

In the case of both completion and English proficiency, the government was already committed to a broad policy direction, and therefore their ability to react was limited. The government response to RBA was therefore defined by the way that it was used as an argument by the senior leadership of the GoR education sector. It helped senior members of the leadership to lock the government itself into the listed priorities; a significant effect in the light of the constantly fluctuating policy context. RBA also supported the government's efforts to coordinate the sector more broadly. The RBA agreement should therefore be understood as helping the government leadership to commit to the targets and to communicate this, by virtue of the funding and the measurement.

6. Has value for money been achieved?

The value-for-money exercise was ground-breaking, in that it applied standard VfM tools to the innovative RBA instrument. The naive interpretation shows that RBA was excellent value for money. However, the assumptions underpinning this result were found to be problematic: for example that RBA caused increases in completion and that these extra exam sitters were no different from other exam sitters. Both of these assumptions may undermine the value for money case for RBA in this particular setting. In terms of the disbursements, we are confident that the performance at P6 and S6 would have happened anyway. Furthermore, funds were disbursed for English improvements which may not have occurred. The reasons for the change in performance that lead to funds disbursed for S3 is less clear, but has not led to a sustained increase in students studying higher grades.

7. What lessons have been learned to inform RBA in Rwanda and elsewhere?

We draw on the collated findings and conclusions to make a series of recommendations:

- New research, available since RBA was agreed, states that agreements should see a good measure as a prerequisite for success. This report concurs wholeheartedly, and finds **RBA should not be used again in the Rwandan education sector unless better measures can be identified.**
- In the specific case of RBA to the Rwandan education sector, there are two possible avenues for future agreements. First, **RBA could use a learning indicator**. Second, **RBA could use English proficiency as the performance measure**.
- **Greater attention should be paid to the political context in which PbR operates, both in future research and agreements.**
- **Efforts should be made to ensure that in future the benefits of understanding and delivering RBA improvements outweigh the costs**. This is not simply about agreeing a tariff, but about agreeing timeframes, working with politics, and, crucially, ensuring financial disbursements are passed on to the relevant sector.
- **Future negotiations should bear in mind that recipient governments have more information on likely improvements than donors**.

1. Introduction and background

Upper Quartile (UQ) in association with the Institute of Policy Analysis and Research – Rwanda (IPAR-Rwanda) are pleased to submit this final report of the evaluation of the Results Based Aid (RBA) in Rwandan Education pilot. This report builds on, and should be read in conjunction with, the year one and two evaluation reports (Upper Quartile, 2014; Upper Quartile, 2015). This report is presented to the Government of Rwanda (GoR) and the UK Department for International Development (DFID).

1.1. Evaluation purpose and scope of work

1.1.1. Purpose of the evaluation

UQ was commissioned to undertake a mixed-methods process and impact evaluation of the RBA pilot in Rwandan education (2012-2014), considering if, how and in what circumstances the RBA pilot contributed to the results envisaged in the Memorandum of Understanding (MoU) agreed between DFID and the GoR.[4] Specifically:

- **The impact evaluation** was to determine "whether or not the additional incentive of results-based aid had any effect on the number of children completing different levels of education compared with what would have been achieved without the provision of this results-based aid" (DFID, 2012: Paragraph 14)
- **The process evaluation** was to consider "the recipient's, and other key actors', response to the RBA, including establishing the processes that led to any increased educational results"5 (DFID, 2012: Paragraph 11)

1.1.2. Scope of the evaluation

The evaluation focused on the RBA pilot. It is noted that RBA is embedded in DFID's wider Rwanda Education Sector Programme (RESP), however this was not intended to be an evaluation of the RESP, and consideration of the RESP is largely in relation to the efficacy of one aid modality (RBA) in comparison to another (Sector Budget Support (SBS)). Given the pilot nature of RBA in this context, the evaluation has a focus on learning lessons that provide feedback on the design of the pilot, and any consequent amendments, as well as contributing to the wider evidence base on RBA in the development sector.

1.1.3. Evaluation questions

During the evaluation inception phase, the evaluation team engaged in an iterative process to determine the evaluation questions. This process involved the lead evaluators, the DFID-Rwanda Education Adviser and the DFID Lead on Payment by Results (PbR). The final evaluation question set offers a balance between areas of interest to DFID-Rwanda and to DFID's central PbR function. It was agreed that the evaluation would seek to answer seven macro-level questions.

- **Question 1: What has been achieved?**
- **Question 2: Has RBA contributed to impact in relation to the envisaged results?**
- **Question 3: What factors have impacted on the achievement of RBA results?**
- **Question 4: How is the RBA approach perceived in Rwandan education?**
- **Question 5: How did government respond to RBA?**
- **Question 6: Has Value for Money been achieved?**

4 'Results' in the MoU are defined in terms of increased **completion** (i.e. additional learners sitting key stage examinations in the sixth grade of primary school (P6) and the third and sixth grades of secondary school (S3 and S6)) and **competency** of teachers in the use of English as the medium of instruction (i.e. additional teachers reaching level B1 proficiency in relation to the Common European Framework for Reference (CEFR) scale.

5 This was to include determining any contribution made by any observed increase in the number of teachers competent to use English as the medium of instruction to any observed increase in the numbers of students completing P6, S3 and S6.

- **Question 7: What lessons have been learned to inform RBA in Rwanda and elsewhere.**

The findings presented in this evaluation report are collated to answer these questions in section 4.

1.1.4. Evaluation timing

The evaluation has taken place over a three year period. The inception phase was completed between April-July 2013 with the evaluation implementation phase running from July 2013 – December 2015.[6] The first and second annual evaluation reports were finalised in March 2014 and 2015 respectively.

1.1.5. Evaluation audiences

The evaluation has a number of target audiences and the findings will be used in different ways by each. It is envisaged that the evaluation will be used by GOR and its agencies (specifically the Ministry of Education (MINEDUC) and the Rwanda Education Board (REB) to aid their understanding of the current situation and the factors impacting completion and teachers' competency in English as the medium of instruction; by GOR and DFID-Rwanda to inform decision making around the design of any future RBA initiatives; and by DFID, HMG and the wider development community to improve understanding of how RBA approaches may be designed and used to maximise the impact of development spend.

1.1.6. Deviations and additions to the original terms of reference

The original TOR (see Appendix 1) remain valid in steering the delivery of the evaluation in terms of its overall purpose and scope of work. There are however a number of deviations from the TOR, agreed in full with DFID, which should be highlighted. These are:

- **Evaluation questions** – The TOR[7] posed a number of evaluation questions. These were amended during the evaluation inception phase (see section 1.1.3 above).
- **The analysis framework** – The TOR states that the service provider should use the current RESP theory of change (TOC) model as a framework to unpack the response of GOR to RBA. Following inception phase discussions, it was agreed with DFID that this was not an appropriate framework for analysis – the RBA pilot was contained within the RESP and there was no RBA specific TOC (either nested within the RESP or stand-alone). It was hence agreed that the evaluation questions would form the overarching framework for the evaluation.
- **Evaluation timing:** The TOR proposed an evaluation completion date of 30 June 2015. A no cost contract extension to December 2015 was agreed with DFID in July 2015. This was to account for delays in previous years in receiving official EMIS[8] data from GOR. This extension meant that the final round of qualitative data collection could be off-set against the quantitative data analysis, allowing qualitative enquiry to explore specific findings of the quantitative analysis in year three.
- **Additions to the TOR**: In year two the evaluation was granted a contract extension to allow more in-depth consideration and modelling of value for money (VfM). The purpose was to explore options for assessing VfM in the context of RBA.[9] The TOR for the VfM assessment are included as Appendix 2. As specified in paragraph 10 of the additional TOR, the evaluation team produced a revised approach paper for evaluating VfM. This paper modified the options outlined in paragraph 8, Table 1 of the TOR.[10] Given the work involved in exploring the cost effectiveness of the RBA pilot, it was proposed by the evaluation team, and agreed by DFID, not to focus on other elements of efficiency and economy as outlined in section 12 of the TOR.

In year three of the evaluation, the approach (a modification of that in previous years) was discussed at length with DFID and formalised in the 2015 Research Approach Paper agreed on 1 July 2015. The

[6] This differs from the timescale proposed in the original TOR. See section 1.1.6 for details.

[7] Indicative questions were posed in Appendix 2 of the original TOR. This is attached as Appendix 1 of this evaluation report.

[8] Education Monitoring Information Systems (EMIS)

[9] As a relatively new aid modality, there is no agreed methodology for assessing VfM. This additional research is contributing to the debate on VfM assessment in the context of RBA.

[10] The Approach Paper (included as Annex 4 of the year two evaluation report (Upper Quartile, 2015)) revised the options in the TOR into two broad models (A and B). Model A covers option 1 in the TOR and model B covers options 2 to 4.

evaluation team and DFID consider that this approach addresses the key requirements of the TOR. This paper, included as Appendix 3, has guided the work of the evaluation team in the year three, and as such supersedes the original TOR.

1.2. Understanding the context

1.2.1. Conceptualising payment by results: the theoretical framework

Broadly defined, payment by results (PbR) is a financing scheme that makes payments dependent on *achieved* results (National Audit Office (NAO), 2015). In aid contexts, PbR is generally arranged in three main ways depending on who the funding recipient is. Results based aid (RBA), the model under consideration in this evaluation, is characterised by government-to-government funding; results based financing (RBF) is government-to-service provider; and development impact bonds are government-to-investors (Bond, 2014). DFID's approach in relation to all three forms of PbR demonstrates two key characteristics:

- First, there is a ***risk transfer*** (from DFID to the partner) as payment depends on a result, rather than an action; and
- Second, payment is contingent on ***independently verified results*** (Clist and Verschoor, 2014).

In the case of the RBA pilot in Rwandan education, the principle of ***recipient discretion*** also features, as the recipient (GOR) was given space to decide how results should be achieved.

There is considerable debate, and multiple views, concerning how PbR modalities may, or should, function in practice to bring about change. One succinct typology, formalised in the writings of the Centre for Global Development, encapsulates these into four major theories of change; (CGD) Perakis and Savedoff (2014). All of these theories draw on the "principal-agent" model from economic theory, whereby a "principal" (the funder) seeks to encourage a particular outcome from the "agent" (the fund recipient) by providing a form of incentive. The four ideas about how PbR creates incentives, posed by Perakis and Savedoff (2014), are:

1. **Pecuniary interest:** That the offer of a payment will create a financial incentive which leads to a behaviour change by the agent – focusing providers on implementation that achieves results (the trigger for disbursal), as opposed to hollow policy reforms;
2. **Attention:** That the performance funding mechanism itself makes results visible and transparent, providing information that supports improved management;
3. **Accountability:** That the focus of both principal and agent on results will improve accountability to constituents or beneficiaries, through the visibility of outcomes**;**
4. **Recipient discretion:** That the focus on results over actions gives agents, who are more attune to the local/ national context and operating environment, greater discretion and autonomy to innovate and adapt their activities.

It is important to note that these four theories are not mutually exclusive – they may exist in any combination and even reinforce each other, depending on the PbR context.

In DFID's strategy documents, PbR is considered as part of the drive to provide value for money; a way to focus DFID's partners on performance while also improving transparency, accountability and innovation (DFID, 2014; NAO, 2015). In this sense, DFID's objectives draw on all of the four major theories of change proposed for PbR.[11]

[11] In relation to the RBA pilot in Rwanda, although no theory of change was defined at the design stage, all four theories are evident from formal documentation relating to the pilot, formal and informal consultations with key informants undertaken as part of this evaluation (see for example the DFID RESP Business Case (DFID, 2011a); Upper Quartile 2014; Upper Quartile, 2015).

It is widely accepted that the potential effectiveness of PbR (common to all four major theories of change) hinges on selection of an appropriate performance measure (Clist and Verschoor 2014; CGD, 2014; NAO, 2015). If payments are not based on measureable performance indicators, which are tightly linked to the desired outcomes, then PbR has the potential to go awry, leading to distortion and generating perverse incentives (see Box 1 for an example).

The issue of the *correct performance indicator*, in relation to the RBA pilot in Rwanda, will be discussed subsequently.

Box 1: Potential for perverse incentives

To illustrate the potential negative effects of PbR, CGD uses a story from colonial-era Delhi:

Administrators, seeking to reduce the number of snakes in the city, offered a bounty for dead snakes. However, this encouraged entrepreneurs to set up snake farms, raising and killing snakes to collect the payments. When the administrators realised this and stopped the programme, the entrepreneurs shut down the farms and released their snakes - leading to an overall rise in snakes in the city.

Because the paid-for results - dead snakes - was not close enough to the desired outcome - fewer snakes – the *principal* actually created a perverse incentive for the *agent* (Talbot and Barder, 2014).

1.2.2. The policy drivers

To understand and learn from the application of PbR in Rwandan education, the pilot should be seen in terms of the broader policy environment. The drivers of PbR in Rwanda emerge from UK and international aid policy trends, rather than the Rwandan context.

Paying for achieved public service outcomes is a concept which dates back to at least the 19th century in the UK (NAO, 2015). In the last decade PbR has experienced a resurgence in British government policy. PbR was initially used by the government for domestic services – applied to health and welfare-to-work programmes – but was also evident in the international development sector (applied to UKAID projects) by 2009 (NAO, 2015). In 2011, the Open Public Services White Paper laid out PbR as a part of cross-government reforms to public service delivery, stating that:

> "…it makes sense to build in an element of payment by results to provide a constant and tough financial incentive for providers to deliver good services throughout the term of the contract. This approach will encourage providers to work more closely with citizens and communities to build services that are both more efficient and qualitatively different" (HM Government, 2011).

Since this White Paper, PbR has played an increasing role in DFID strategy; demonstrated by the increasing number of international aid projects with a PbR component (NAO, 2015) and by DFID's 2014 Strategy for PbR which stated that 71% of centrally issued contracts in the preceding 12 months had a performance-based element (DFID, 2014, p.12)

In recent years, under a variety of nomenclature, other donors have also been making more use of performance-based payment schemes (Perakis and Savedoff, 2014). While DFID has played a role and introduced PbR into the development lexicon, this international trend is also part of the legacy of the Paris Declaration. Recognising that aid could produce better impacts, this forum placed *managing for results* and *mutual accountability* as two of the five fundamental principles of effective aid (OECD, 2005/2008). Hugely influential, this contributed to the growing international trend for "Cash on Delivery" Aid (COD Aid), Results-Based Finance (RBF), Results-Based Aid (RBA) and other PbR-related schemes. This includes significant programmes by influential aid organisations like the World Bank and development thought leaders like the CGD (DFID, 2014).

However, despite this British and international trend, the evidence base in support of PbR is still very limited. This is noted in numerous documents, reports and strategies. For example, in their report on PbR in the British government, the NAO specifically notes that "there *is little existing evidence of effectiveness*" (NAO, 2015). Similarly, CGD notes that RBA and cash on delivery is "*largely untested*," (Perakis and Savedoff, 2015b) and DFID's own PbR strategy admits that the evidence base is "*still at an early stage*" (DFID, 2014).

Partly because of this lack of evidence, the NAO notes in their review that "PbR is not suited to all public services" with "attendant costs and risks that government has often underestimated." The NAO go on to caution that "applied inappropriately there is a risk that either service quality or value for money may be undermined". As a result, the NAO recommends that "commissioners should justify their selection of PbR over alternatives" (NAO, 2015).

Given the lack of evidence, the RBA pilot is best understood as a way of generating evidence of whether this new modality can offer solutions to the perceived problems in the UK policy context.

1.2.3. Payment by Results in Rwanda

The RBA pilot in Rwanda formed part of DFID's £74.98 million GBP Rwanda Education Sector Programme (RESP). The RESP was embedded in GOR's Education Sector Strategic Plan (ESSP); the sector wide approach (SWAp) guiding all education sector planning and spending in Rwanda.

The RBA agreement was stipulated in the Memorandum of Understanding (MOU) between the GOR and DFID agreed in October 2012. The MOU specified an overall budget for the RBA pilot up to a maximum of nine million GBP, with an expected disbursement schedule up to a maximum of three million GBP per year for three years from 2013 (with the first payment being allocated to results achieved in 2012). It was later agreed that any shortfall in a given payment year could be rolled over to subsequent years.

The RBA agreement was intended by DFID to drive change in the education sector in ways that were agreed government priorities.[12] As such, RBA results were also priority results of the ESSP. A 'hands-off' approach was adopted by DFID with GOR free to achieve results and utilise RBA funds as they saw fit. RBA incentives were applied to improvements in completion at key stages of education and improvements in teacher proficiency in English as the medium of instruction. The details of the agreement are set out in Box 2.

Box 2: The RBA Agreement in Rwanda

The final GOR-DFID agreement (DFID and GOR, 2012) sets out four results to be incentivised via RBA. It stipulates that RBA payments would be effected as follows:

- Where for a given province and gender there is an improvement above the 2011 baseline for the first time, a payment of £50 GBP will be applied for P6 and S6.
- Where for a given province and gender there is an improvement above the 2011 baseline for the first time, a payment of £100 GBP will be applied for S3.
- Where performance at P6, S3 and S6 is above the 2011 baseline but not for the first time, a payment of £10 GBP will be applied.

  For all of the above payments, for any given province-gender pair, payment will be made for improvements with any drops discarded.
- In 2015 DFID will pay the GOR £50 GBP per additional teacher competent to use English as the medium of instruction. This will be based on a baseline assessment in 2012 and a follow-up assessment conducted in 2014. Any payment due will be made in 2015 based on independently verified results and subject to available funds within the £9 million GBP three year ceiling.

All of the above stated payments were subject to the independent verification of the results.[13]

The DFID RESP Business Case states that the Results Compact (i.e. the RBA agreement) will:

[12] In line with the Paris Declaration principle of ***Ownership*** i.e. countries set their own strategies for development (OECD, 2005)

[13] Independent verification was undertaken by HEART.

> "reward a year-on-year increase in learning achievement of girls and boys at key stages in their schooling. This component will ensure the focus of MINEDUC is on improving learning outcomes, not just increasing enrolment" (DFID, 2011a).

This statement is illustrative of DFID's desire to enhance learning outcomes via the RBA agreement. However, completion in itself is not a measure of learning. A 'completer' is defined as a student who sits the final exam; achievement in that exam was not considered. Records documenting the RBA negotiations show that a measure of educational quality and learning outcomes, while desirable, was not considered feasible as an RBA indicator at the time.[14] The selection of 'completion' as the indicator for the pilot can therefore be considered a pragmatic choice. This is not a criticism, indeed completion as a proxy indicator in these circumstances has been supported by key proponents of PbR modalities. Savedoff (2010) said:

> "the COD Aid team consulted with experts and was able to identify an outcome measure – the number of assessed completers – that was reasonably simple to understand and either solved or mitigated a number of foreseeable problems…By including a test, the effect of the program on quality of schooling becomes visible and can be monitored" (Savedoff, 2010)

However, given the implications of the indicator for the potential effectiveness of PbR schemes (see section 1.2.1 previously), this is an important issue to raise and one to which we will return subsequently.

By contrast, proficiency in English as the medium of instruction was included as an RBA indicator at the insistence of GOR, specifically REB.[15] The year two evaluation report outlines the reasons for this – highlighting that English language proficiency was a major policy focus for GOR at that time, as well as an area that posed significant challenges (Upper Quartile, 2015). Evaluation evidence (Upper Quartile, 2014; Upper Quartile, 2015) shows that DFID were not convinced on the inclusion of the language indicator, but consented to ensure the pilot could go ahead.

1.3. The evaluation findings to date

The year one evaluation drew limited conclusions on the impact and effectiveness of RBA in the context of Rwandan education and highlighted mixed findings in relation to the reaction of government. The headline finding was that the RBA pilot did not make a SIGNIFICANT contribution to the observed increase in completion in 2012. This was established through analysis of results from two econometric models and corroborated by the qualitative fieldwork.

In year two, the evaluation concluded that GOR had achieved SIGNIFICANTLY above trend completion at the S3 level, linked to increased access, increased retention (in particular a remarkable improvement in converting S3 enrolees to S3 completers in 2013) and slightly improved transition (an increase in the number of S2 enrolees who went on to S3 in 2012). This positive improvement was not replicated at P6 and S6 where completion was either negligibly or SIGNIFICANTLY below trend. Qualitative research in year two was not able to pin-point the mechanism by which the increase in S3 completion had come about, [16] concluding that while RBA may have successfully reinforced GOR efforts, GOR policy implementation more widely, as opposed to RBA specifically was the likely cause of the increase.

[14] At that time there was no agreed quality standard (assessments were norm referenced, not criterion referenced) and there was no baseline for an agreed standard. Furthermore, there were multiple possible measurement tools and, at that point, there was no annual measurement of learning outcomes undertaken in Rwanda. Inclusion of learning outcomes as the RBA indicator would have required a representative primary research sample study of schools to be undertaken annually.

[15] This is apparent from the review of documents charting the RBA design process and from consultation with KIs within DFID.

[16] The sequencing of the research in year two, which had been agreed with DFID, meant that qualitative research was conducted before completion of the econometric modelling. As such the qualitative enquiry explored the RBA response and processes for improving completion more generally as opposed to a specific focus on the S3 'bump'. In an attempt to find an answer, the evaluation team conducted additional qualitative research in February 2015 where the data was put directly to the GOR education Senior Management Team (SMT). Still no explanation was identified.

2. Methodology

2.1. Methodological approach

The approach to the evaluation of RBA in Rwandan education stems from a *realist perspective*, rooted in the recognition that outcomes are affected positively and negatively by the real world context in which they occur (Pawson and Tilley, 1997; Stern et al, 2012). Realist evaluation recognises the complexity of interventions in the social world and the difficulty of isolating the impact of a single intervention, seeking instead to explore what works, for whom, in what circumstances and why. The evaluation of RBA in Rwanda has taken place over three years. In line with the realist approach, the evaluation methods remained flexible, evolving to meet the needs of the study and the client group.

2.2. The evaluation framework in year three

In years one and two of the evaluation, the seven macro-level evaluation questions (see section 1.1.3) formed the overarching framework for research, analysis and reporting. In year three, given the changing needs of DFID for evidence on RBA, the complexity of RBA as a modality, and the subtleties emerging in the research findings on RBA in the Rwandan context, it was considered that an amended approach was required in order to provide the nuanced evidence sought.

Following discussions with DFID and the evaluation reference group, it was agreed that the evaluation questions would be retained, for reasons of consistency and practicality of considering the available evaluation evidence base in the round, whilst using the evidence generated to date to more tightly define the research avenues to be pursued in year three.

On the basis of the combined evidence base (i.e. the evaluation in year one and year two, and the econometric analysis in year three[17]), three specific issues that were of interest to DFID and which required further 'unpacking' were identified. From these, the evaluation team formulated three research hypotheses. The overarching evaluation framework in year three is centred on these hypotheses; using complimentary methods to test the hypothesis and address the evaluation questions therein. The three research hypotheses are shown in Table 1.

Table 1: Research hypotheses in year three

1	**Null hypothesis (H0)**	RBA did not contribute to the observed increase in S3 completion in 2013
	Alt. hypothesis (H1)	RBA did contribute to the observed increase in S3 completion in 2013
2*	**Null hypothesis (H0)**	RBA did not contribute to the observed increase in S6 completion in 2014
	Alt. hypothesis (H1)	RBA did contribute to the observed increase in S6 completion in 2014
3**	**Null hypothesis (H0)**	RBA did not contribute to the observed increase in teachers' proficiency in English language
	Alt. hypothesis (H1)	RBA did contribute to the observed increase in teachers' proficiency in English language

*This is discussed in section 3.1.2 in relation to the econometric findings in year three.
**This is discussed in section 3.2.1 in relation to the results of the follow-up survey of teachers' English language proficiency.

The research hypotheses are discussed in more detail in the 2015 Research Approach Paper, included as Appendix 3.

2.3. Methods

Table 2 summaries the methods used in each year of the evaluation. The table demonstrates how each of the methods complement one another and meet the needs of the research.

Table 2: Research method summary

[17] Sequencing of evaluation methods in year three was altered to allow completion of the initial econometric analysis prior to detailed development of the qualitative approach.

	Method and lead evaluator	Description	Strengths	Weaknesses	Years		
					1	2	3
Quantitative	**Econometric modelling (P. Clist)**	Modelling exercise drawing on national level secondary data to identify any effect in RBA-years over and above what may have been expected.	• Possible to identify SIGNIFICANT change. • Objectivity. • Reliability.	• Cannot answer the 'how' and 'why' questions. • Cannot control for all confounding factors.	✓	✓	✓
Quantitative	**Value for money (VfM) analysis (J. Holden)**	Considers cost-effectiveness of RBA relative to not providing RBA using national level secondary data and standard VfM practice to construct two counterfactuals.	• Objective basis for considering cost-effectiveness of RBA. • Contribution to the theoretical debate on RBA. • Recognised approach allows comparison with other interventions.	• Built on number of contestable assumptions.		✓	✓
Qualitative	**Desk-based document review (B. Whitty)**	Desk-based policy and literature review to situate the RBA pilot within an understanding of the prevailing political and economic processes in Rwanda.	• Combined with KIIs to explore incentives, relationships, distribution and contestation of power. • Considers how RBA functions as an incentive.	• Desk based analysis may fail to uncover subtle and ongoing change that is not formally documented.	✓	✓	✓
Qualitative	**National level KIIs (B. Whitty)**	Semi-structured interviews with national level stakeholders to explore the response of GOR to RBA (latterly with a focus on the year three research hypotheses).	• Insight into stakeholder understanding of RBA and reasoning around incentives.	• Subjectivity. • Researcher presence may influence findings.	✓	✓	✓
Qualitative	**District and school-based KIIs and FGDs (B. Whitty and IPAR)**	Semi-structured KIIs and FGDs with district/ sector education officials and school Principals* to explore issues and trends in completion and English language proficiency (latterly with a focus on the year three research hypotheses).	• Possible to uncover complexities of what works, where, why and how.	• Subjectivity. • Context dependent. • Small sample limits 'generalisability' . • Researcher presence may influence findings.	✓	✓	✓

* In years one and two of the evaluation, English language mentors, teachers, parents and students were consulted.

2.3.1. Quantitative methods

2.3.1.1 Econometric modelling

The impact evaluation is centred around an econometric modelling exercise to identify what has been achieved in terms of completion, and the factors impacting on completion, with the aim of isolating any observable effect of the RBA pilot.

As the RBA pilot in Rwanda was implemented nationally, it was not possible to establish treatment and control groups to identify any impact of RBA in an experimental evaluation. The econometric modelling essentially established an artificial counterfactual, modelling trends in completion at key stages of

education and controlling for various influencing factors to understand what might have been expected in the absence of RBA.[18] OUT-OF-SAMPLE predictions were used to test the accuracy of the models. The econometric modelling in year three builds on the finding from years one and two. The data limitations persist (section 2.4), but are eased slightly by the addition of another year of data.

Two econometric models using publicly available data have been developed for the purpose of the evaluation. The two models act as a check on each other since a conclusion supported by both provides a stronger evidence base than a conclusion based on just one set of assumptions.

- **Model 1** builds a counterfactual by relying upon time trends and recent district performance in completion to project into the future. Model 1 does not use district characteristics, but instead relies on district DUMMIES, a time-trend and year-DUMMIES. The advantage of Model 1 is that it requires very little data and can thus exploit a longer time series;
- **Model 2**, which was spilt into two parts (Model 2a and Model 2b) for the year two evaluation, takes a different approach. Both Models 2a and b use as much information on district and time differences as possible, including additional information such as the number of classrooms and teachers in each district. However, Model 2a uses only information that is available on an annual basis, which allows for district level fixed effects and clustering of the standard errors. Model 2b uses all available information, including data for which there is only one OBSERVATION per district. For example, annual data on population by district is not available. Model 2a ignores differences in population by district, with all static cross-district information being 'soaked up' by the district DUMMIES. Model 2b provides an estimate of the effect of the district-level differences that are observed.

Table 3: The econometric models – a summary

	Model 1	Model 2a	Model 2b
DEPENDENT VARIABLE	Exam sitters by district, level and year.	Exam sitters by district, level and year.	Exam sitters by district, level and year.
INDEPENDENT VARIABLES	District and year DUMMIES; time trend.	All data that is available by district and year.	All available data.
Advantages	Allows standard errors to be clustered and unobserved fixed effects to be controlled for. Allows cross-gender correlation to be controlled for.	Allows standard errors to be clustered and unobserved fixed effects to be controlled for.	Provides an estimate of the effect of all relevant characteristics. Allows cross-gender correlation to be controlled for.
Disadvantages	May 'over fit' the relationship and assumes historical trends continue in perpetuity.	Does not exploit cross-gender correlation or provide rich policy-relevant information.	May bias standard errors downwards. Assumes homoscedasticity.[19]

2.3.1.2 Value for Money (VfM) assessment

As a relatively new aid modality with a very limited evidence base, DFID is particularly interested to understand the value for money (VfM) offered by RBA, particularly relative to other aid modalities. With no established methodology for completing this assessment, in year two of the evaluation UQ was awarded a contract extension to allow for more in-depth consideration and modelling of VfM in relation to the RBA pilot in Rwandan education. The TOR for this assessment are included as Appendix 2.

The approach to the VfM assessment is informed by standard practice in assessing VfM in the education sector (DFID 2011c, 2011d, 2014), the fundamental principles of RBA (Clist and Verschoor,

[18] It is not possible to fully control for all factors, especially national-level changes which coincide with RBA. Qualitative research helps understand the influence (or not) of RBA.

[19] Homoscedasticity means that the size of errors are constant across the model e.g. that errors are not much larger in districts with higher completion rates.

2014) and more pragmatic concerns regarding the feasibility of conducting the analysis. Justification of the approach taken and the assumptions made are explicit in the VfM discussion at section 3.4.

2.3.2. Qualitative methods

The econometric modelling and VfM exercises explored changes in the numbers of completers at key stages of education, the cost, and potential returns, of the investment that has achieved this. However, the econometric modelling and VfM exercises are unable to conclusively establish the cause of any increase in completion or the extent to which these may be products of RBA, as opposed to other factors in the external environment. This was the aim of the qualitative research.

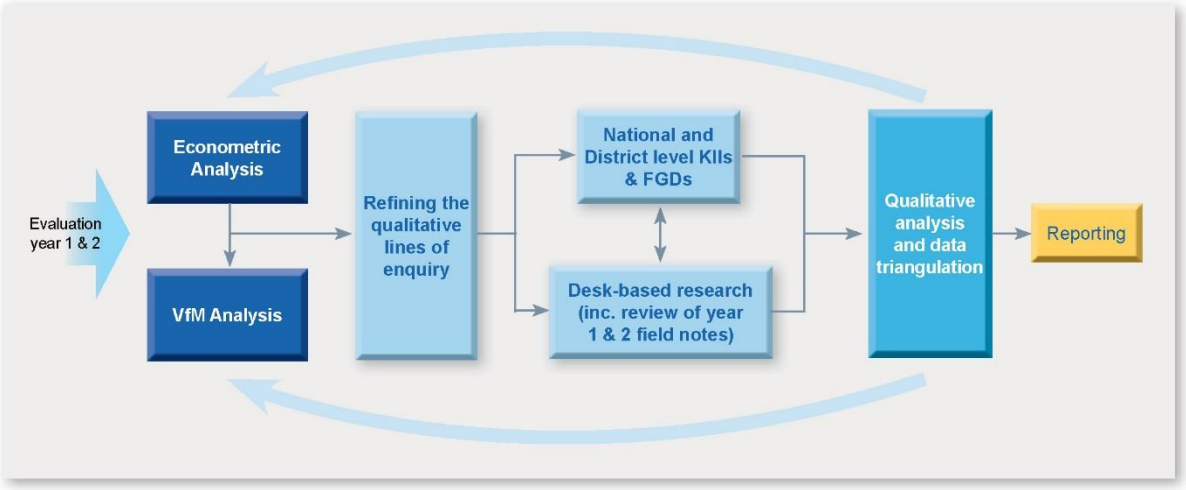
In year three the qualitative research was delayed to allow completion of the econometric modelling exercise before designing and undertaking the final round of qualitative fieldwork. This meant that the qualitative research could be focused to explore specific findings of the econometric modelling in year two (a 'loose-end' from the previous annual evaluation report) as well as exploring interesting features of the 2014 completion data and the follow-up survey of English language proficiency; the three research avenues outlined in the 2015 Research Approach Paper (see Appendix 3). The qualitative research in the final year of the evaluation involved three related components:

- **Desk-based review**: A rapid review of policy and literature to provide any necessary updates to the position set out in the comprehensive political economy analysis (PEA) completed in year one of the evaluation. Desk-based research in year three also involved a review of field-notes from years one and two of the evaluation with a focus on identifying any evidence that may inform the three research hypotheses outlined in Table 1.
- **National-level Key Informant Interviews (KIIs):** Semi-structured interviews were conducted with national level stakeholders to explore the research hypotheses set out in Table 1. Guided by the existing evidence base (from previous years of the evaluation, the econometric and VfM analyses in year three) KIIs explored the SIGNIFICANT findings of the econometric model, the range of factors that may explain these findings and any possible contribution of RBA. Twenty-one KIIs were conducted with representatives from GOR (MINEDUC/ MINECOFIN), REB, international donors and NGOs working in the education sector. A list of national level consultees in year three, and the rationale for their inclusion, is in included as Appendix 4. The research tool is at Appendix 5.
- **Local level KIIs and Focus Group Discussions (FGDs):** Semi-structured KIIs were conducted with (formerly titled before a reorganisation) District Education Officers (DEOs) and Sector Education Officers (SEOs) in three districts. FGDs were also conducted with school Principals in each of these districts. For reasons of continuity and exploring existing findings in a more nuanced way, the districts selected in year three were the same ones visited in year two. KIIs and FGDs used open-questioning to explore district-level completion data before probing, on the basis of the existing evidence, to understand the mechanisms by which change has been brought about. The local level research programme was smaller in the final year of the evaluation than in previous years. The evaluation in year three was never intended to include district or school-level research, as such a pragmatic approach in line with available resources was required. The 2015 Research Approach paper (Appendix 3) sets out the rationale for this approach. Potential limitations are discussed in section 2.4. The composition of district-level research participants in year three, and the rationale for their inclusion, is in included as Appendix 4. The research tools are at Appendices 6 and 7.

2.3.3. Analysis and triangulation

In year three of the evaluation a mixed-method sequential analysis process was applied – with the quantitative econometric and VfM analyses shaping the qualitative lines of enquiry to be followed up through desk-based and field research. This process is illustrated in Figure 1.

Figure 1: Mixed-method triangulation and analysis process



Following completion of the in-country field based research, the evaluation team came together for a 'round table' analysis session (held in London on 7 September 2015). At this session the lead evaluator for each research method presented their findings (for the last year of the pilot (2014) and in the round). For each of the three research hypotheses and evaluation questions, the full evaluation team worked together to collate the research findings and formulate conclusions. Table 4 shows how the research methods combine to address the evaluation questions.[20]

Table 4: Addressing the evaluation questions

Evaluation Questions	Econometric modelling	VfM	Context and PEA	National KIIs	District / school KIIs/ FGDs
1. What has been achieved?	✓				
2. Has RBA contributed to impact in relation to the envisaged results?	✓		✓	✓	✓
3. What factors have impacted on the achievement of RBA results?	✓		✓	✓	✓
4. How is the RBA approach perceived in Rwandan education?				✓	✓
5. How did government respond to RBA?			✓	✓	✓
6. Has value for money been achieved?	✓	✓			
7. What lessons have been learned to inform RBA in Rwanda and elsewhere?	✓	✓	✓	✓	✓

This collaborative process placed the quantitative findings in context and resulted in some revision to the interpretation of quantitative evidence; specifically the interpretation of data from the value for money analysis. The full evaluation team are in agreement about these revisions. The analysis session also highlighted a small number of areas where qualitative evidence warranted follow-up and additional exploration of quantitative data; this was specifically in relation to data presented in the RBA verification report (HEART, 2015). This is discussed further in section 3.1.3.2.

[20] An expanded version of this table showing the evidence sources which have contributed to answering each evaluation question is provided in the 2015 Research Approach paper (see Appendix 3).

2.4. Limitations of the evaluation

The evaluation team has identified various limitations of the evaluation approach. Many of these, especially those associated with the econometric model, persist from year one. Limitations of the evaluation are set out below.

2.4.1. Limitations of the econometric evidence

The main limitations for the econometric model come from the availability and quality of the data. The incentive to over report enrolment data in the Rwandan school system is well known (Sandefur and Glassman, 2015), and if the bias is different in different years then measurement error would affect the accuracy of the results. The econometric report (Appendix 8) presents OUT-OF-SAMPLE tests, which show that the models perform well given (quite severe) data limitations and quality. In practice this means that there are wide confidence intervals, as the model cannot be more precise in forecasting future trends. The main weakness relates to annual changes, as the sample size here is small. For example, the ability to control for cohort effects is limited, as we only observe a small number of transitions from (e.g.) P6 to S3.

2.4.2. Limitations of the Value for Money evidence

The VfM analysis is in part based on the results of the econometric modelling. It is therefore subject to the same limitations. In particular, test B2 of the VfM approach – the VfM of RBA compared to other forms of aid – is built directly from the econometric modelling. The detailed assumptions of the VfM model, and associated limitations, are set out in section 3.4 and in detail in Appendix 9.

2.4.3. Limitations of the qualitative evidence

In terms of limitations of the qualitative work, the politically sensitive nature of the interviews means that care has to be taken in interpreting responses at face value. Where possible, findings have been triangulated with previous years' work and with available data from other research strands – fortunately this year the research was able to build on the findings of the quantitative work, which has helped to focus the questions and increase confidence in the findings. Among other limitations are:

- **The availability of key people within GOR** presented the national and district level research with considerable challenges. At the national level, interviews were often hard to schedule and were frequently rescheduled at short notice. This exacerbated the resource constraints presented by the relatively short time for the research, and limited the space for triangulation of the findings.
- **Turnover of staff in senior positions within MINEDUC and REB,** meant that certain key personnel were not available – notably, the former Permanent Secretary of MINEDUC. This limits the degree to which it has been possible to confirm and triangulate the information provided.
- **Resource constraints meant that it was only possible to conduct interviews in three districts.** The relatively small number of interviewees and the limitations on the number of districts mean that confidence in the conclusions should be modified accordingly.

2.5. Research ethics

All research carried out by Upper Quartile and IPAR-Rwanda was conducted in accordance with DFID Ethics Principals for Research and Evaluation (DFID, 2011b)

All informants consulted in the course of this three-year evaluation consented to participate.[21] With the exception of senior MINEDUC and REB officials, who granted permission to be identified, no individual is named in the report and the names of schools and districts have been removed.

[21] In the case of children and young people consulted in years one and two of the research, a responsible adult was also required to grant consent. Only children aged 10 years and over were included as informants.

All IPAR researchers who conducted primary field-based research in local communities and schools have completed training in research ethics. Ethical considerations were reviewed before fieldwork commenced.

All members of the evaluation team are fully independent. IPAR-Rwanda is an independent think tank based in Kigali; Upper Quartile is an independent consulting firm based in the United Kingdom.

3. The evaluation findings

This chapter presents a discussion of the evaluation findings in 2014 and over the three-year pilot period as a whole. To set the discussion in context, Table 5 presents the overall final payment breakdown from the RBA pilot in Rwandan education. The headline from Table 5 is that over the three years of the pilot, £5,440,590 GBP was disbursed from a possible maximum of £9 million GBP; approximately 60% of available funds have been disbursed as RBA (although it is noted that the £9 million GBP ceiling figure was not a target).

Table 5: Payment Overview

Level	Total payment 2013 (based on 2012 completion)	Total payment 2014 (based on 2013 completion)	Total payment 2015 (based on 2014 completion)	Total
P6	£562,950	£90,380	£122,610	£775,940
S3	£364,200	£1,477,320	£90,330	£1,931,850
S6	£228,900	£315,720	£1,010,080	£1,554,700
English	-	-	-	£1,178,100*
Total	£1,156,050	£1,883,420	£1,223,020	£5,440,590

* The English language payment was recommended in 2015 for 2014 improvements on the 2012 baseline.
Source: HEART, 2015.

In sections 3.1 and 3.2 that follow, the evaluation findings are discussed separately for the two RBA indicators of completion and English language respectively.

3.1. Completion

3.1.1. RBA Payments

FINDING 1	Trends in completion are impressive for both secondary school levels, and virtually flat for primary school.

FINDING 2	Increases in completion are at least partially affected by cohort effects which were determined before the RBA agreement was active.

Of the circa. £5.4 million GBP overall RBA payment, 78% (circa. £4.2 million GBP) was for improvements in completion. To recap the tariff structure:

- Where for a given province and gender there was an improvement above the 2011 baseline for the first time, a payment of £50 GBP was applied for P6 and S6.
- Where for a given province and gender there was an improvement above the 2011 baseline for the first time, a payment of £100 GBP was applied for S3.
- Where performance at P6, S3 and S6 was above the 2011 baseline but not for the first time, a payment of £10 GBP was applied.
- For any given province-gender pair, payment was made for improvements with any drops discarded.

Table 6 presents the absolute number of exam sitters by grade, gender and year for the baseline in 2011 and for the three years in which RBA was operational. It also shows the change in completion 2013-14 (the final year of the pilot).[22] The final row shows the percentage change in exam sitting by gender-grade pair over the lifetime of the programme.

Table 6: Exam sitters by grade, gender and year, 2011-14

	P6		S3		S6	
Pilot year	Male	Female	Male	Female	Male	Female
2011	70,548	84,406	38,043	39,377	24,535	22,023
2012	74,877	91,276	37,754	42,836	24,237	26,065
2013	73,552	89,542	44,227	49,505	26,689	27,102
2014	70,782	86,341	40,246	45,845	34,978	34,305
2013-14 Trend	↓	↓	↓	↓	↑	↑
Change 2013-14	-4%	-4%	-9%	-7%	+31%	+26%
2011-14 Trend	→	↑	↑	↑	↑	↑
Change 2011-14	0%	+2%	+6%	+16%	+43%	+56%

Note that S6 includes those taking vocational and teacher training courses (TVET and TTC).

Source: HEART, 2015.

The headlines from Table 6 are that:

- **Over the course of the pilot as a whole, there has been an increase at the higher grades in the numbers of students sitting exams, and stalling at P6.** Specifically:
  - P6 saw annual fluctuations, but in the final year of the pilot overall exam completion was only 1% higher than in the baseline year of 2011.
  - Completion at S3 increased over the period, most dramatically in 2013.
  - Completion at S6 increased to the greatest extent over the period, driven by a dramatic increase in 2014.
  - In terms of gender differences, 2014 saw more boys sit S6 exams than girls, but there is essentially gender parity over the last few years. P6 and S3 are less equal, and changes over the four year period (from the baseline in 2011) have increased that difference, meaning more females sit exams than males (by 22% and 14% respectively).
- **In 2014 annual changes were negative for P6 and S3. The only positive changes observed were at the S6 level.** It is also noted that:
  - The final year of the pilot (2014) can be considered a poor year for P6 completion, with the absolute number of exam sitters down 5,971 on the previous year. The 2015 payment for P6 performance – £122,610 – illustrates the effect of payments being made on province-gender pairs, and disregarding drops. Much of this will be due to natural fluctuations in different provinces.
  - Similarly, 2014 was a disappointing year for S3 completion (although this is reflective of the fact that 2013 was an exceptionally good year due, in part, to the large cohort of students).[23]

[22] These figures are presented for previous years of the pilot in the first and second annual evaluation reports (Upper Quartile, 2014; Upper Quartile, 2015) and can be easily derived from Table 6 if required.

[23] The year two evaluation findings devoted considerable effort to examining the large increase in S3 completion from 2012-2013, attributing around half of the increase to cohort effects.

- In 2014 a large jump in the absolute numbers of students sitting the S6 exam is evident; an increase of around 50% from a steady base.

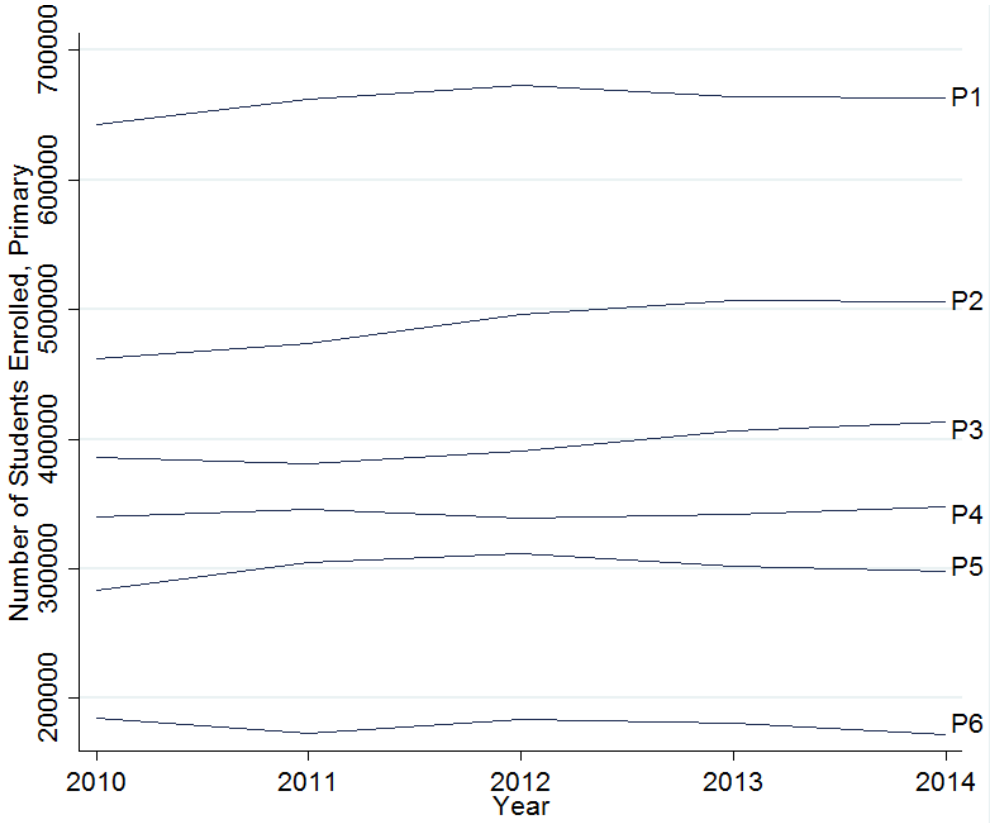
To place the above stated observations in context, these trends should be understood in terms of a relatively stable number of enrolees into a school system that gradually 'leaks' students in each grade, as they drop-out.

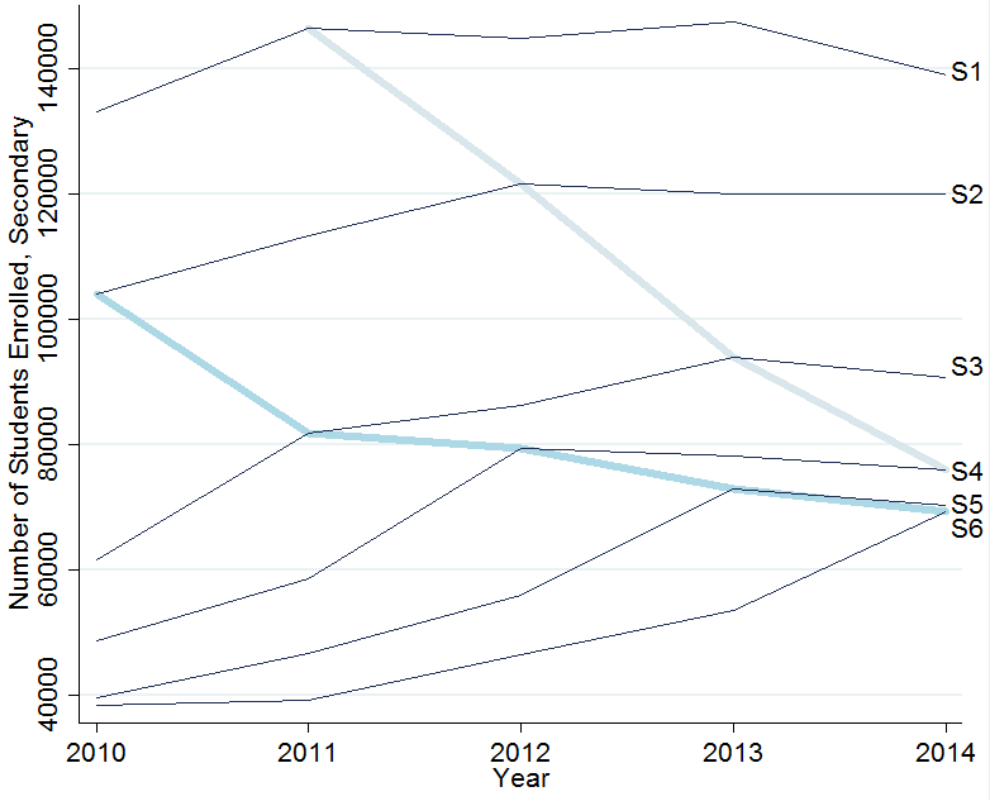
This is illustrated in Figure 2, which shows the number of students enrolled in each grade over the course of the RBA pilot period. Enrolments in P1 are consistently over 600,000, but less than 200,000 enrol in the final year of primary school. This process is mirrored in every grade, with lower drop-outs and repetition guaranteeing higher exam sitting. In essence, the number of enrolees in a subsequent year provides the framework that limits or enables current improvements.

In this context it is unsurprising that cohort effects have driven many of the effects on completion; where a large cohort of students is enrolled in a given year this is associated with a large number of enrolees in the next year and, subsequently, completers.

The blue lines in Figure 2 show that high S1 enrolment in 2011 filtered through to high S3 enrolment in 2013 (a year with a significant 'bump' in completion figures (see Table 6)). Similarly, high S6 enrolment in 2014 can be traced back to high S2 enrolment in 2010.

Figure 2: Enrolment by year and grade





Source: GOR EMIS Data – various years

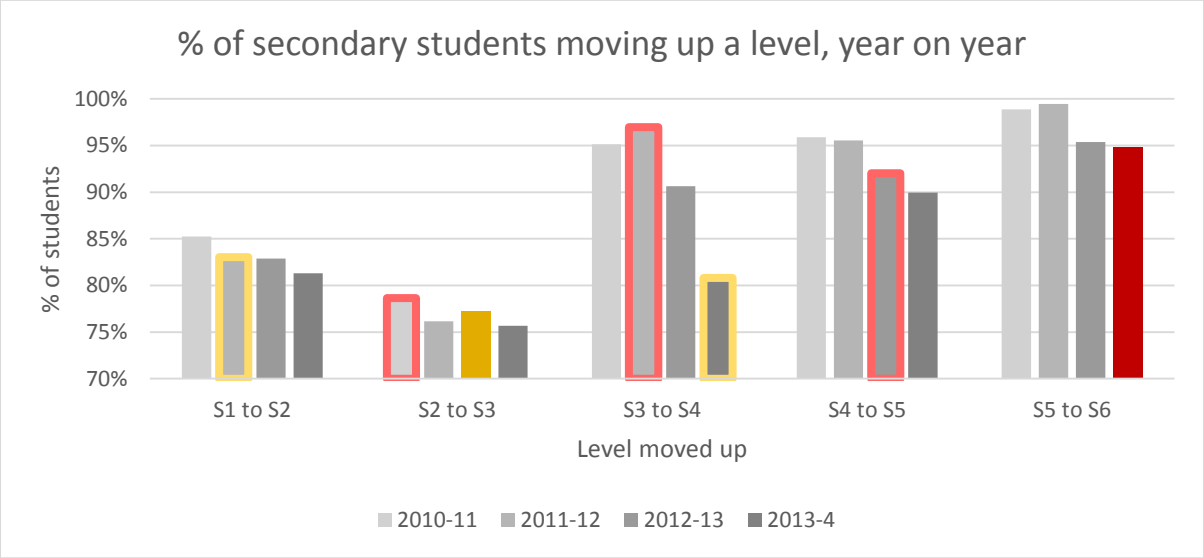
The effects of cohort size can be seen more clearly in Figure 3, which plots the transition level between different grades for the life of the RBA pilot. For example, the left most bar shows the number of students that enrolled in S2 in 2011 as a percentage of students that enrolled in S1 in 2010 (e.g. the percentage of students who moved up a year).

The large cohort of S3 exam sitters in 2013 can be tracked through its transition rate at different levels by following the yellow bars, with the other cohort of interest (i.e. those taking the S6 exam in 2014) shown in red. The cohorts are traced through the years with the red and yellow outlines while the S3 and S6 'bump' years are solid. These cohorts are also highlighted in Table 6.

The graph illustrates that in the first year of the RBA pilot, this cohort did not see an abnormally high number of students transferring from S1 to S2. The transition from S2 to S3 was similarly comparable to other years: all bars in the second group are between 75% and 80%. This implies that the high number of S3 exam sitters in 2013 was related to the number of students enrolled in this cohort before the baseline year. A concerning development for this cohort is shown in Figure 3 however, as the cohort does see a large drop in transition from S3 to S4 in 2013-14.

The red bars track the cohort that ultimately took the S6 exam in 2014. Again, the transition rates are comparable to other years, implying that the high number of exam sitters was heavily influenced by the number of students enrolled before the RBA MOU was agreed.

Figure 3: Percentage of students moving up a level, year on year



GOR EMIS Data – various years

The trends in absolute numbers of students completing exams is impressive at the secondary school level, with virtually flat performance for primary school. Prima facie, much of the improvement appears to be related to cohort size effects which were determined before the RBA agreement was signed.

3.1.2. A significant change?

FINDING 3	There is no evidence of a sustained increase in completion as a result of RBA during the years that RBA was operational at any level. While the absolute numbers of completers has increased over the period, the vast majority of performance is well within the expectation of what would have happened in the absence of RBA.

FINDING 4	**SIGNIFICANT** differences from expected performance (i.e. the level of completion predicted by the econometric models in the absence of RBA) were as likely to be below trend as above.

FINDING 5	S3 performance in 2013 and S6 performance in 2014 was **SIGNIFICANTLY** above trend, and merits further investigation.

Absolute completion figures, the basis upon which the RBA was paid, do not in themselves show if completion rates (as a percentage of enrollers) are rising or the extent to which there may or may not be an RBA effect. To explore these questions we turn first to the findings of the econometric modelling exercise. Table 7 summarises results from the econometric models highlighting the grades and years where the quantitative analysis requires further enquiry; instances where completion is SIGNIFICANTLY different than what may have been expected in the absence of intervention.

Table 7 contains the final results from four different approaches to the econometric analysis, highlighting SIGNIFICANT effects by year.

Table 7: Summary of results from econometric models

Level	P6	S3	S6
Summary statistics	Unambiguously, 2014 was a poor year for P6 completion, with 4% decrease from 2013.	2014 was much worse than 2013 for S3 completion, but this appears to be related to 2013 being an exceptionally good year partly due to cohort-size effects.	2014 saw a remarkable 50% increase in Regular S6 completion. This appears to be a cohort-size effect.
Model 1	RBA was negative where SIGNIFICANT, and never positive.	RBA was INSIGNIFICANT, but possibly better for girls.	RBA was negative where SIGNIFICANT, and possibly better for girls.
Model 2a	The baseline was SIGNIFICANTLY below trend. Even against this, RBA was occasionally SIGNIFICANT and negative (2014 for boys)	RBA-years were generally on trend. 2013 was a SIGNIFICANTLY good year for girls, but this could have been a cohort-size effect.	RBA-years were generally on trend, but for boys 2012 was SIGNIFICANTLY low and 2014 SIGNIFICANTLY high. This could be a cohort size effect.
Model 2b	The baseline was SIGNIFICANTLY below trend. Even against this, RBA was occasionally SIGNIFICANT and negative (2014 for boys and girls, 2013 for boys).	In RBA-years, 2013 was SIGNIFICANTLY above trend for both genders and 2012 SIGNIFICANTLY below for boys. Cohort effects are difficult to control for.	Both genders were SIGNIFICANTLY below trend in 2012. In 2013 for girls and 2014 for boys, completion was above trend. Cohort effects are difficult to control for.

Model 1 and the summary statistics are the least sophisticated models as they do not control for factors that will influence the numbers sitting an exam. Models 2a and 2b control for factors such as the number of teachers in a given district, with model 2b controlling for all available information. Further detail on the different models and data on OUT-OF-SAMPLE performance (a common test of accuracy), can be found in Appendix 8.

All models concur that there was no SIGNIFICANT positive effect of RBA at the P6 level, with occasionally SIGNIFICANTLY negative performance. The 2011 baseline is also found to be quite low, which may have artificially inflated the naively measured performance. The detailed econometric report (Appendix 8) illustrates the effect of the 2011 baseline, as opposed to a scenario where the higher (in absolute terms) 2010 baseline had been used. If the 2010 baseline had been used, the 2012 payment for P6 performance would have been 40% lower. **The econometric models concur that P6 performance was *not* above trend in years where RBA was active, and often below it.**

Analysis of completion at S3 and S6 gives a more mixed picture. **There are two cases in which a SIGNIFICANT and positive effect was found: S3 in 2013 and S6 in 2014.** In both cases these were one-off observations with an above trend number of students sitting the relevant exam; there is no

evidence of any sustained and positive effect of RBA at any level. In fact, in the year after the S3 'bump', the number of students enrolled in S4 actually fell in absolute terms. Worryingly, this suggests that there is unlikely to be an S6 (senior secondary) 'bump' in 2016 as may have been expected given performance at the junior secondary level in 2013. Both of these peaks in completion (S3 in 2013 and S6 in 2014) are worthy of further investigation by means of qualitative enquiry.

Given the above, can it be said that there was a SIGNIFICANT change in completion during the years RBA was active? While the absolute numbers of completers has clearly increased over the period, the vast majority of performance is well within the expectation of what would have happened in the absence of RBA. Where the numbers of completers was SIGNIFICANTLY outside of expectation, it was as likely to be below trend as it was to be above trend. That said, the two cases of above trend performance that are noted are worthy of further investigation to understand if, and to what extent, RBA may have contributed.

3.1.3. The explanatory factors for change

This discussion begins by exploring the S6 increase in 2014 as the explanation for this appears straightforward.

3.1.3.1 Increased completion at S6 in 2014

FINDING 6	Above trend completion at S6 in 2014 is the result of a cohort effect. The mechanism by which this has been bought about is the policy shift by GOR to introduce fee-free education. RBA did not contribute to the increase.

While viewed as a whole, S6 completion during the years that RBA was applied was below trend, completion in 2014 was actually SIGNIFICANTLY above trend with the final year of the pilot seeing S6 exam sitters jump by around 50%. Although the negative effect in 2012 (noted in Table 7) may lead this 2014 positive effect to be overstated, it is clear that the 2014 increase should still be investigated.

Further quantitative analysis implies that the effect is likely due to a large cohort – the high number of S6 exam sitters in 2014 can be clearly and logically linked to high S6 enrolment in 2014. This, in turn, can be traced back over the years to a high S2 enrolment in 2010 (see Box 3).

Box 3: The quantitative analysis explained

Referring to Figure 2, the dark blue band traces a large cohort from enrolment in S2 2010 through to the point of sitting the S6 (Senior Secondary) exam in 2014. It shows that as this group progresses through grades, large enrolment is seen in each year as compared to previous and subsequent years. In the context of this large cohort, the increase in S6 exam sitters in 2014 is consistent with previous years. Over the course of the evaluation period, progression from S5-S6 has been within the range 95-99%; with a 95% progression rate, 2014 is actually at the lower side (see Figure 3). **The quantitative analysis therefore supports the assertion that *above trend* completion at S6 in 2014 is a cohort-size effect.**

To strengthen the evidence base, the evaluation sought to validate this finding through qualitative enquiry. Overwhelming consistency in responses was noted with the increase linked by every interviewee, both at the national level and in the districts, to the introduction of fee-free schooling; particularly the introduction of 9-year Basic Education (9YBE) in 2009 (and subsequently 12-year Basic Education (12YBE) in 2012). It is noted that the large S2 cohort observed in 2010 (the first year for which the evaluation team has historical data) would have enrolled in S1 in 2009; corresponding to the introduction of 9YBE. Supporting this finding, a key informant from REB commented:

> *"In S6, the reasons were linked to the introduction of 9YBE. In 2003 we introduced Universal Primary Education – UPE. Then in 2009, we introduced 9YBE, and they continued, and then in*

2012 we introduced 12YBE. We got the Commonwealth Award in 2012 for our 9YBE. So they [the students] ran all the way through: it was a cohort effect." ***[KI interview, REB (2015)]***

In this case it is apparent that GOR policy changes have widened the availability of fee-free schooling, triggering increased enrolment and completion. The qualitative research at national and district level reinforces the quantitative analysis and confirms the null hypothesis that RBA did not contribute to the observed increase in S6 completion in 2014.

3.1.3.2 Increased completion at S3 in 2013

FINDING 7	Three explanatory factors contributed to the increase in completion in S3 in 2013: (1) lagged uptake of 9YBE; (2) community and school-led efforts to combat dropouts; and (3) a Directive requiring students in senior secondary who may have progressed without sitting the S3 exam, to return and sit the exam.

The above trend increase in completion at S3 in 2013 poses a greater puzzle. This 'bump' in completion was identified in the econometric modelling undertaken for the year two evaluation. No credible explanation for the increase could be found by either the initial round of qualitative research (August 2014), or subsequent focused follow-up research in February 2015.

The year two report, noted 'additional' completion in the range of 13,300 exam sitters. Further analysis, suggested 6,000 of this was attributable to a cohort effect, 1,300 to improved transition from S2-S3 and 6,000 to a large increase in percentage of S3 enrolees who went on to sit the S3 exam (Upper Quartile, 2015).

The econometric analysis in year three of the evaluation (see Appendix 8) adds an additional year of data to the calculations, which has modestly increased the ability of the models to control for cohort-size effects (a stated limitation of the evaluation approach).[24] This has resulted in a revised estimate of 'additional' completers at S3 in 2013 – a slight downward revision. Nonetheless, the increase remains SIGNIFICANT and in need of explanation. Box 4 outlines the quantitative findings.

Box 4: The quantitative analysis explained

The year two evaluation report concluded that some of the additional completion at S3 in 2013 could be attributed to cohort-size effects. Additional data from the year three analysis supports this conclusion. Referring again to Figure 2 it is possible to trace another large cohort as they progress through secondary school – this group is illustrated by the thick light blue line in Figure 2. For this cohort, high S1 enrolment in 2011 filters through to high S3 enrolment (and subsequent completion) in 2013. A cohort-size effect in 2013 is also supported by analysis at district level (see detailed econometric analysis in Appendix 8) with lower 2014 completion figures suggestive of 'regression to the mean' following the outlier year (2013). However, unlike the 2014 S6 cohort (discussed in Box 3), the timing of the large cohort in this case is not aligned neatly with a policy shift in Rwanda's implementation of fee-free education (i.e. 2003, 2009, 2012). Further, the 2014 drop-off in enrolment at S4 is not consistent with the explanation that cohort size is the only explanatory factor here.

As Box 4 shows, the question raised in the year two evaluation report – why did so many S3 enrollers go on to sit the S3 exam in 2013? – remains unexplained by the quantitative analysis.

Qualitative enquiry in year three of the evaluation set out explicitly to understand this change. At both the national and district levels the evaluation team invited interviewees, through open questioning followed by specific probing questions, to explain both the increase in S3 completion in 2013 and the subsequent drop in 2014.[25]

[24] As it allows another view of the extent to which lagged completion of P6 affects S3 completion.

[25] Relevant completion figures were made available to interviewees as a discussion prompt

REB and MINEDUC officials were unable to give definite explanations for this increase, and most were unaware of it. When asked, most speculated that the increase was due to the lagged effect of 9YBE and/ or ongoing school-building efforts, but were unable to illuminate why – in that case – there was such a substantial increase in 2013 (as compared to 2012) followed by a subsequent drop in 2014. In prompted discussion some suggested that efforts by the centre to reduce repetition and drop-out were responsible (see Box 5).

Box 5: Comments and quotes from national level officials

"*I'd propose that the first reason is that we took in so many children who were not going to school, so that now they are exhausted from that source. It was a big effort for the first years, and now almost everyone is enrolled in school – so we have taken everyone off the streets, the streets are exhausted. I think the second thing is the accuracy in statistics. Records are getting better. Some head teachers we know are exaggerating the number of children they have."* **[KII, REB official, 2015]**

This quote refers to differences between the EMIS data and the REB data on exam sitters. Both are considered measures of completion – the EMIS data is the one the government typically refers to, and which the REB official here is referencing; the REB the measure on which RBA disbursement is conditional. The EMIS data is generally considered to be of poor quality – until recently it was filled out by head teachers. The REB examinations data – used in RBA – is verified and considered to be much better quality (see Sanderfur and Glassman, 2015).

Officials from REB and MINEDUC (2015) commented on: Construction of schools to accommodate growing numbers of students.

An official from REB attributed the increase to: The effects of Town Hall meetings conducted by key REB and MINEDUC officials which sought to communicate the government priorities on completion to district and school-level officials (see also Upper Quartile, 2014 for more details of Town Hall meetings) **[KII REB (2015)]**

"I'm just thinking District put more effort to sensitise Head Teachers to reduce the drop-out, [ ] the Districts put more effort to encourage schools not to repeat…" **[KII, MINEDUC official, 2015]**

"There were challenges around the distance from the home to school – we chose to offer [different school] combinations for those who preferred to travel farther; to address the risk of drop-out." **[KII, REB official, 2015]**

As in year three of the evaluation, no clear or consistent picture emerged at the national level, and there is no official Government of Rwanda position on the figures. By contrast, a very clear and consistent position emerged through qualitative interviews in the districts, which highlighted three potential explanations.

The first is linked to the lagged effect of the introduction of 9YBE spurring wider efforts of the education sector to accommodate an increasing number of students, with a consequent increase in faith in the system which has driven enrolment (and completion). This explanation from staff 'on-the-ground' is consistent with the perceptions of some national level officials (when probed to explain observations on S3 completion) and is suggestive of a general trend within Rwanda; albeit the evaluation sample at district level is considered only illustrative (and not representative). Officials tied this to the ongoing programme of school construction, where schools were build not only to accommodate increased volumes of pupils but also distributed in such a way as to reduce the travelling time. Quotes from district level officials included in Box 6:

Box 6: Quotes from district- and school-level officials #1

"The 9YBE increased the number of S3 completers because not all nine year basic education pupils started in 2009, some started in 2010 while others started in 2011: across the years these numbers kept rising." ***[FGD, Principal District A (2015)]***

"there has been a change of mind set in parents from the previous years, and since 2009 to 2013 the system was stable for students to join 9YBE schools. Another thing is that from 2009 to 2013, more schools were built and so the increase in the number of students [occurred]." ***[FGD, Principal District B (2015)]***

"…before parents and students couldn't trust the schools but as years passed and schools were equipped with materials, skilled young teachers; all that attracted the students and so they enrolled at a high rate." ***[FGD, Principal District B (2015)]***

"Today the system of 9YBE is now firm and dependable as years pass from 2011 when the first candidates of 9YBE sat for national exams; parents have trusted the system with their children after the results of the past national exams showed that even students from 9YBE can pass it. So this increased the number of joining students, to the extent of children moving boarding schools to 9YBE schools and this increased the S3 completers in 2013." ***[FGD, Principal District C (2015)]***

The second explanation, cited relatively consistently (if not in every interview at the district level), emphasised efforts to reduce drop-out and repetition at the sector and school-level. Captured in Box 7, these citations also reflect findings from the year two evaluation where community efforts to reduce drop-out emerged consistently as factors perceived to influence completion. It is however noted that other arguments posed in the year two evaluation (the influence of enhanced options for TVET, the impact of the School-Feeding Programme) were raised only in passing in this years' research, and only sporadically.[26] These are not therefore considered influential in this overarching evaluation report.

Box 7: Quotes from district- and school-level officials #2

We worked together with other authorities like Community education officers (Abajyanama b'uburezi) in reducing the dropout rate where students from poor families were supported as others were also encouraged and motivated along with their parents, and so children came back to schools" **[KII, SEO District A, 2015]**

"There is a government policy of focusing in reducing dropout where local leaders motivate children to be at school, this have increased the number of students where you find no children just staying at home without going to school. Even there was an emphasis of following up all students who registered to sit for national examination, e.g. we used to have students who could register and don't sit for the exams, but now who ever registered is now found to sit for the national exams." **[Focus Group Discussion, Principal District A, 2015]**

"But as years passed, there came other levels to follow up education e.g. the Sector Education Officers that work closely with schools and check in their numbers of students in classes, where they could know how they promote and repeat students and other movements in schools as they advise the principals, check any errors found. This level of SEOs helped a lot in the influence to reduce repetition rate in schools. SEOs started in 2012 in January and I think their impact was experienced in 2013 in many things as well as the increase of completing students. **[KII, former DEO District B, 2015]**

Finally, the third explanation posed by district level interviewees is linked to a directive issued by REB in 2013. This Directive, (which was not mentioned in national level interviews and has not been alluded to in previous rounds of the evaluation research) is reported as having re-articulated the pre-existing requirement that every student who wished to sit the national examination in S6 must first possess a certificate for S3. District level findings from year three of the evaluation suggested that students who had gone to private school – and may in many cases not have acquired the relevant certificate – re-entered the schooling system in 2013 to allow their further progression. Interviewees

[26] Other key factors identified in the year two report – such as the automatic promotion – may be explanatory factors for broader questions about grade progression than the narrower focus of S3 in 2013.

suggested that this prompted these students to sit for the S3 exam in 2013. Quotes from district level officials are included in Box 8:

Box 8: Quotes from district – and school-level officials #3

"there was a case of students who have escaped the S3 examination in the years before 2013 who were going to sit for S6 examination. In 2013 REB gave directives that no student will sit for S6 without a certificate of S3, many of the students returned to S3 to sit for the examination" ***[Focus Group Discussion, Principals District B (2015)]***

"There was an instruction in 2013 calling out all students who hadn't sat their S3 exams to do so. A number of students had proceeded to S4 – S6 (A level) due to either fear of failure or the fact that there was no requirement for a certificate to advance to the next level for some private schools." ***[Focus Group Discussion, Principals District C (2015)]***

The consistency in reporting in relation to this third potential explanation was striking at the district level; all but one of the nine interviews with district level education officials or SEOs mentioned this in unprompted discussion as a key reason explaining the increase in S3 completion in 2013, and it was raised in each of the three districts visited by the evaluation team. However, quantifying this accurately is difficult, as the following quotes from the district and school level interviews show:

Box 9: Quotes from district- and school-level officials #4

I don't think this made the number rise to thousands and thousands of students' increase." ***[KII, DEO District C (2015)]".***

"at last [students in senior secondary] all came back in 2013 for the S3 certificates which they had to sit for the National examination of S6. In [name of school] they had 300 candidates to sit for S3 national exams." [27] ***[Focus Group Discussion, Principals, District B (2015)]***

"In 2013 I remember that there was a case of some students, who came back from upper classes to sit for S3 exams and that time I also had more than 80 additional students and are the ones I knew, meaning there are others who went to sit for S3 exams in other districts and that's why there is an increase in 2013. " ***[KII, DEO, District A (2015)]***

"At my school I had 5 students adding to those who were at my school registered for the S3 exams yet they were in S6. … Though many other students were forgiven and sat for S6 examinations due to being late to register for S3 so they could sit for it in the following year" ***[Focus Group Discussion, Principals, District C (2015)]****,*

Although consistently cited as an explanatory factor, and on occasion identified as the "main" factor, it is difficult to identify the scale and the prevalence of this factor; and this may have varied across districts, sectors and schools. As a general indication, extrapolating from the figures of the third and fourth quotes (namely, 80 children per district or 5 per school), suggests either 2,400 students (80 children in each of 30 districts) or 7,605 (5 in each of 1521 schools). This represents, respectively, 16% and 51% of the total payout for S3 at 2013, and may be some indication of the order of influence.

At both the district and the school level this REB directive was also cited consistently as an explanation for the subsequent drop in S3 completion in 2014, since those who had previously missed the exam had mostly now acquired the necessary certificate. When approached with this information, a senior official in REB suggested that this may indeed be the explanatory factor.

As previously stated, RBA payments are linked to the independent verification of results. The evaluation team has therefore followed-up with the verification team to see if this potential explanation could hold

[27] For a crude illustration, according to the EMIS data for 2014, there are 43,944 S3 children in 1,521 schools, giving an average of almost 29 per school.

true. The head of the verification team reported that their methodology would not have identified that this kind of examinee was from a different year to S3.

In conclusion, there are three explanatory factors for the increase in S3 in 2013: the first is the lagged uptake of 9YBE, linked to the growing trust in the quality of the education and the school construction programme; the second is the community and school-led efforts to reduce drop-out; and the third is the directive from REB to the effect that senior secondary students without a certificate from S3 must come back and sit the S3 exam.

Of these, RBA is not considered as a potential influencer in factor one or three. The first was part of a long-term GOR programme that was initiated long before RBA, while the last is an administrative re-emphasis of an existing rule and any increase cannot be taken to represent a 'new' S3 sitter.

The potential contribution of RBA in relation to the remaining factor – community and school-led efforts to reduce drop-out – is discussed subsequently in section 3.3.

3.2. English language

3.2.1. RBA Payments

The evaluation in years one and two looked only briefly at teachers' proficiency in English as the medium of instruction (EMI) – recapping the findings of the baseline survey undertaken by the British Council in 2012 and presenting qualitative findings from national and district level stakeholders with regards to perceived progress and processes associated with EMI implementation. It was not possible in previous years to comment on achievements in relation to language proficiency or any possible effect of RBA.

In 2012 the British Council undertook a survey involving a sample of 557 teachers to ascertain baseline levels of English proficiency in relation to the six levels of the Common European Framework for Reference (CEFR) (Council for Europe, undated). Results of the 2012 baseline survey, illustrated in Table 8, showed that the vast majority of teachers (96.8%) possessed only a basic level of English language proficiency; having been assessed at level A0-A2.[28]

The follow-up survey, used as the basis for independent verification of RBA results and calculation of any amount payable via RBA, was undertaken in November/ December 2014. The follow-up survey showed a marked improvement, with 43.4% of teachers assessed at the B1 intermediate level.[29] This is compared to 2.9% in 2012.

Table 8: Overall CEFR Results

Level	2012 baseline	2014 follow up
A0	0.2%	0
A1	41.3%	9.2%
A2	55.5%	41.8%
B1	2.9%	43.4%
B2	0.2%	5.4%
C1	0	0.2%

Source: British Council Survey of Teachers' Proficiency in English, 2012; 2014 Endline Assessment of English Language Proficiency of School Teachers in Rwanda, 2015.

Following independent verification of these test results, the verification team concluded that DFID should make a payment of £1,178,100 GBP to GOR for the English language component of the RBA agreement (HEART, 2015).

[28] **A1: Breakthrough or beginner** - *Understands and uses familiar everyday expressions and very basic phrases;* **A2: Waystage or elementary** - *Understands sentences and frequently used expressions. Communicates routine tasks requiring a simple and direct exchange of information.*

[29] **B1: Intermediate or threshold** - *Understands the main points of standard input on familiar matters regularly encountered in work, school, leisure, etc. Can produce simple connected text on topics which are familiar.*

3.2.2. A *real* change?

FINDING 8	The baseline and endline tests, and associated conditions, of English language proficiency were not comparable to the extent that it is not known whether there was any increase in proficiency.

Unfortunately, while the results presented in Table 8 and discussed in section 3.2.1 appear impressive, there are a large number of concerns regarding the data. Taken together, this means it is impossible to draw conclusions about RBA's impact on any improvement in English language proficiency. Indeed, the results mean that it is not possible to tell if there was, in fact, any improvement in the English language proficiency of teachers at all. In short, the 2012 and 2014 results are from different exams, taken under different conditions, and are therefore cannot be compared. There are three main differences which mean that the baseline and endline are not comparable:

- The 2012 baseline used the **Aptis General** test, while the 2014 follow-up used **Aptis for Teachers** (a test which includes vocabulary and scenarios that are more likely to be familiar to teachers).
- In 2014, three hours of training were given immediately before the test. This included information such as test taking techniques and a practice exam in order to mitigate issues that may have adversely affected results in 2012 (such as lack of familiarity with computers and difficulty understanding the British accent used in the test).[30]
- In 2014, a much greater effort was aimed at reducing stress and anxiety, as in 2012 it was felt some teachers were concerned over the consequences of the exam to the extent that it affected performance.

Each of these issues are discussed by the British Council (2014, pp.12-13) in the report of the endline assessment but none are mentioned in the report summary (British Council, 2015). Each of these differences are expected, by the British Council, to have made the test a more accurate measure of teacher's English, but also to have positively affected the results. As such, it is not possible to know how much of the measured increase (shown in Table 8) is due to any true improvement in the English proficiency of teachers, and how much is due to the use of a different test under different conditions.

In fairness to the British Council, the Terms of Reference for the 2014 follow-up assessment appear to have placed greater emphasis on implementation of an accurate test, as opposed to a comparable one. Qualitative research undertaken for the evaluation in year three highlighted that changes to the test methodology were made in response to recommendations following the baseline assessment. Key actors from both the donor community and GOR reported that the initial test was viewed as unfair and that they wished to have a diagnostic that would be of use as a data-set for policy-making; fairness of the test and utility of the results was prioritised over a consistent test of improvement (see Box 10).

Box 10: Quotes regarding the test of English language proficiency

"*Everyone wanted it to be a useful test, and to be valuable outside the function of the RBA results – and there had been a number of suggestions from the baseline about changes, quite outside its function for RBA, about making the test more meaningful about capturing the capability in English. So it seemed a bit funny not to take these on board – so there was a balance between practicality and the needs of RBA*." **[KI, International Community (2015)]**

"*The change was because there were recommendations at the end of the baseline to do something that would be more fair – so DFID built that into the TORs, they agreed with REB that the test would take on board these recommendations*" **[KI, International Community (2015)]**

[30] The use of a British accent in the original test, questions which were out of context, teachers' lack of familiarity with computers and their general anxiety over the possibility that they might lose their jobs led many to say the original test was unfair. The revised test provided more contextualisation, removed scenarios possibly unfamiliar to Rwandan teachers, doubled the test familiarisation period, and changed many test components from online to pen-and-paper.

The broader lessons and consequences regarding this are discussed further in section 3.3.1. However, it does appear that a substantial amount of RBA funds were disbursed on the basis of data that was not able to substantiate whether there was any actual change in the English proficiency of teachers.

3.2.3. The explanatory factors for change

In spite of the issues above, for the sake of completeness (and to fulfil the requirement of the TOR to examine the response of GOR to RBA), the section that follows discusses the mechanisms by which any change in language proficiency may have been brought about. This is a necessary precursor to section 3.3 in which we consider any influence of RBA in these mechanisms.

The switch to English as the Medium of Instruction (EMI) in 2009 was a high level decision made to connect Rwanda into the East African Community, a political decision which came from the highest level of GOR. The scale of the challenge presented by this shift was considerable. The evidence-base for the impact of such a switch in other African countries emphasises the negative effects (see for example Negash, 2011). Among education sector stakeholders in Rwanda, this move was met with concern; the evaluation in each year has highlighted perceived and ongoing resistance from some teachers (particularly among older teachers from a Francophone teaching background [**KIIs and FGDs with Principals (2015); KIIs and FGDs with Principals, English Language Mentors and Teachers (2014)]**). In 2010, senior officials in the education sector successfully persuaded the GOR leadership to alter its policy, delaying the age at which EMI was introduced until Primary 4.

There are several policy points in particular that evidence from the district-level identifies as explaining any changes that may have happened. However, as stated above, it is impossible to say whether any positive or negative changes in language proficiency have actually occurred.

First, the main government intervention in relation to English language was the School Based Mentoring Programme. This emerged from prior efforts focussed on centralised mass-tuition of the teaching profession in the Christmas holiday period (2009-11). In 2011, it was decided that this was too expensive and unlikely to be effective. At this point the newly created REB decided to move to an English language mentoring programme as the core response to issues of language proficiency – there was "*no framework and no vision on the part of Government of Rwanda beyond the mentorship programme*" **[KII, International Community (2015)].**

In late 2012 a major conference was held in Gisane – co-hosted by DFID – which sought to develop a strategy for the SBMP and secure the buy-in of major international actors to SBMP priorities. The SBMP Strategy was used to coordinate the efforts of development partners but, although some adaptations were made (e.g. the introduction of Senior Mentors), the broad approach was already firmly in place and was to continue in spite of the scepticism of many within the sector.

The evaluation evidence reveals tensions and mixed views on the mentorship programme; the programme was contested in policy discussions and the Technical Working Groups addressing teacher training and proficiency. Alternative strategies[31] were formulated (see Upper Quartile, 2015 for details) but not adopted during the period of the MOU on RBA. Following a recent change of leadership in REB, and after considerable challenges in implementation and recruiting mentors, the SBMP has been significantly restructured with the non-Rwandan mentors being sent home.

Box 11: Quotes regarding the School Based Mentoring Programme

"Mentors made great contributions and were productive, although they had no scheduled time to mentor teachers in their busy schedules, mentors helped those who had some basics in English. Elderly teachers who resisted and kept using French didn't learn anything." **[FGD, Principals District B (2015)]**

"There were some mentors who didn't work at all because they had two schools and were managed by REB... Another problem was that most of the mentors were foreigners and couldn't help in

[31] National strategy for Teachers' In-Service Training to support English as the Medium of Instruction (EMI).

language understanding … not all mentors did badly because some really did a great work in improving teachers' English like at our school we had scheduled the mentor in the evening hours after classes... Generally the REB mentors failed, I remember there were other mentors once came to help from IEE NGO who were supervised and well organised and contributed a big thing in English" **[FGD, Principals, District A (2015)]**

"*Mentorship started in 2012/2013, and only teachers who were determined to learn English worked hard for themselves and improved in English but the programme of mentorship didn't do anything big in English improvement."* **[FGD, Principals, District C (2015)]**

"*It was very unclear if SBMP was the cause. There's so much variability, and there are real doubts about the quality of mentors; it's not clear what the value added is of the SBMP. But there has been a positive effect; and there have been other moves towards continuous professional development."* **[KII, international community, (2015)].**

The second key factor identified by school and district officials is the training of better teachers. As one sector-level interviewee observed, the increased numbers of Rwandan graduates emerging with good English has made a difference:

> *"Today in our sector we have many teachers who studied in English and they are graduates from KIE, e.g. some of our teachers studied with mentors in same schools, meaning our teachers are good in English".* **[KII, SEO District A, 2015)]**

The emergence of this group has been used by management to incentivise existing teachers, as two interviewees in REB's senior leadership indicated; their communication to the established teaching profession is that they will face competition from these new teachers, and will be expected to deliver.

The third aspect is the general management communication effort to the teaching profession. As part of the effort to focus on English, REB and MINEDUC have insisted that EMI and the importance of English proficiency would be crucial to teachers' promotion and training opportunities.

Box 12: Quotes regarding communications to the teaching profession

"The REB leadership went round the districts monitoring and evaluating what was happening. By 2015, teachers knew that in the future for their promotion prospects English proficiency would be a major factor – they knew that this was going to be key, they knew that future training prospects would depend on it. The teachers were sensitised in these things – and they also were aware that ICT skills were going to be just as important, it wasn't just English language." **[KII, REB official, (2015)]**

"We keep saying that in any Town Hall, it is possible to make the change, it is possible to do this [to teach] in English." **[KII, REB official, (2015)]**

The fourth and final aspect identified was the coordination of the international community behind their plan. Some interviewees noted benefits from other forms of assistance and the delivery of other materials, particularly through the L3 project run by EDC.[32]

Box 13: Quotes regarding donor activity

"In 2014, L3 programme was introduced and materials to use in teaching like books and telephone to use in teaching mathematics, Kinyarwanda and English by listening where provided. These materials are computers with CDs where teachers trained to use those computers and follow it in their teaching while kids or students are listing" **[KII, SEO District A (2015)].**

[32] L3 is the 'Literacy, Language and Learning Project' funded by USAID and implemented by the Education Development Centre (EDC) in collaboration with MINEDUC/ REB.

"Generally English improved in 2014 in both teachers and students due to the text books, listening materials by EDC USAID gave and the MINEDUC/ REB soft English materials given in 2013." **[KII, SEO District C (2015)]**

Officials in REB noted that in the conference in Gisane in 2012 they sought to coordinate the sector to support their plans on English. Different development partners were able to do this to differing extents, depending on their own internal priorities and the limitations that they had on their own budget allocations.

3.3. RBA as a factor contributing to change

Section 3.1 and 3.2 presented the evaluation findings in relation to the indicators of completion and English language proficiency – identifying what has been achieved and the processes/ mechanisms by which change has/ is perceived to have been brought about.

In this section we consider the contribution (if any) of RBA to the process. This analysis uses, as its starting point, the dominant theories around how PbR (including RBA) will play out in the development sector, outlined initially in section 1.2.1.

Box 14 suggests that these four theories can be collapsed into two broad categories: first **the way in which RBA creates an incentive effect on the leadership of the education sector** (attached either to the financial reward for results achieved or to the creation of an external accountability mechanism); and second the **effect that it has on policy formulation** (by increasing discretion and providing an evidence-focus in the form of increased attention on data)**.** In the discussion that follows, the contribution of RBA will be understood through these two overarching categories.

Box 14: Categorising RBA Theories of change

Creating incentive effects	
Pecuniary interest	That financial incentives directly stimulate actions (such as changes in policy, practice and messaging) to achieve results and receive funding.
Accountability	That the requirement for independent verification of results (as the trigger for fund disbursal) creates a process for measuring achievement which in turn permits other accountability forces to exert a constructive force on the sector. It is noted that CGD's writings on this subject focus largely on RBA as a driver for 'bottom-up' accountability "to constituents" (i.e. civil society holding the government to account).
Effects on policy processes	
Recipient discretion	That it gives discretion to government, by proposing an outcome and allowing freedom to achieve it. This is considered to improve policy-making, ensuring greater compatibility with, and responsiveness to, local knowledge, building local capacity and driving innovation and adaption
Attention	That performance funding (with a requirement for independent verification) makes results visible in a way that improves management by enhancing the evidence base upon which to make reforms, adapt policies and programmes

Sections 3.3.1 and 3.3.2 consider the effect of RBA in relation to completion and English language proficiency respectively.

3.3.1. Completion

FINDING 9	Improvements in completion were weakly to moderately incentivised, taking in particular the form of a concerted communication effort to reduce drop-out through community- and school-level efforts. This communication mechanism was incentivised by RBA moderately and was relatively efficacious, although the RBA argument had a very weak influence on the communication efforts around community and school-level dropouts. This contributed to the rise in S3 sitters in 2013 caused by the community and school-focus.

As discussed, over the course of the RBA pilot DFID has disbursed £4,262,490 GBP in funds linked to improved completion, in terms of absolute numbers, at three key stages of education. The evaluation's econometric modelling identified two points at which completion is considered 'additional (i.e. over and above what would have been expected in the absence of intervention) and where it is therefore possible that RBA may have been a contributing factor. These were S6 completion in 2014 and S3 completion in 2013. Section 3.1.3.1 conclusively links the first of these to a cohort-size effect as a result of policy change by GOR. There is no evidence that RBA has played a role and the evaluation considers that the same effect would have been observed in the absence of RBA.

The *additional* completion observed at S3 in 2013 is, however, a more complex picture. The previous discussion suggests three factors that, in combination, may have brought about the observed change:

- **Explicit communication by REB requiring that all those looking to sitting an upper secondary exam (S6) must first be in possession of a junior secondary (S3) certificate.** This appears to be the 'smoking gun' of the year three evaluation; the elusive explanatory factor that was not identified in year two. Qualitative research suggests that this was an administrative directive issued by REB leadership. Senior secondary students returning to sit an exam they had earlier skipped are not the kind of exam-sitter anticipated or desired by the RBA agreement. While there is no evidence to suggest that it may have been motivated by a desire to 'game' the RBA agreement, there is also no way to rule that possibility out. Indications from interviews suggest that the S3 increase was a surprise, and therefore that the possible effect of this particular directive may have been unanticipated. Since it is neither desired, nor is it possible to establish a link to intent, the evaluation will not pursue this further as a possible RBA-linked driver.
- **Increased enrolment (and exam sitters) linked to a lagged take-up of 9YBE.** Qualitative enquiry links this to prior GOR decisions and commitments around 9YBE which were bearing fruit by 2013 as capacity and confidence in the sector caught up with policy implementation. Once again there is no suggestion that RBA had any effect.
- **Community and school-level efforts to reduce drop-out**. If RBA has contributed to additional completion, this is the most likely mechanism. It is very hard to quantify the efficacy of this factor against the other potential explanatory factors.

How, then, could RBA be said to influence community and school-level efforts to reduce drop-out? In line with the structure at section 3.3.1, the following discussion considers how RBA worked as an incentive on the senior leadership (3.3.2.1), and as an influence on policy (3.3.2.2.). Once again this analysis proposes slightly different tracks within these categories.

3.3.1.1 RBA in creating incentives

In relation to suggestions that RBA *incentivises* the sector, the often stated term 'Rwanda is a results-driven country' is reiterated.[33] Completion was a government priority entrenched within the ESSP. The importance of combating drop-outs and repetition was understood widely at the local government and school level.

As with English proficiency, interviewees in the senior leadership of MINEDUC/ REB suggested that they were incentivised by the knowledge that money would come in. It is very difficult to verify the effect of this in any quantitative sense (given the strong existing focus in this area and the fact that no major

[33] This was highlighted in all years of the evaluation (see also Upper Quartile, 2014 and Upper Quartile, 2015)

change in approach to achieving completion has been observed since the implementation of RBA). If it incentivised, it did so very generally by adding an additional impetus to the pre-existing drive for completion (a psychological incentive of sorts). RBA may therefore be seen to reinforce existing accountability mechanisms. Any contribution of RBA sat alongside other motivations to achieve results on completion (e.g. those linked to internal management incentives and results targets set by the government hierarchy).

Several points serve to weaken the incentive effect of RBA in relation to completion. They are demonstrated in the evaluation evidence base across the three years of the RBA pilot. These are:

- That there were already strong incentives for education sector stakeholders to take completion seriously. Completion was a GOR priority set out in successive ESSPs and targets on dropout/ repetition (key factors impacting completion) were embedded through the Imihigo processes (see Upper Quartile, 2015 for a more detailed discussion).
- As stated in relation to English language, the government was treating RBA as an experimental modality – they were not sure how it was supposed to operate, and they accepted it with a view to explore its meaning. Further, any incentive effect was to some extent based on a misunderstanding about the disbursal of funds. The incentive was therefore vague and ill-defined.[34]
- The completion target was defined differently within the government in comparison to the RBA MOU – the key ESSP target concerned, not sitting the exam but, enrolment in the following year.[35]
- In contrast to the English language baseline, data on completion was not new and there is no evidence that it has been used more widely to drive improved performance (demonstrated by the lack of knowledge of and inability to explain data anomalies such as the 'bump' in S3 completion in 2013).

Accordingly, we conclude on the basis of the available information that the incentivisation effect was **weak-moderate** (depending on whether they had realised how DFID would disburse), and insofar as it had any effect, it acted to reinforce existing internal management accountability structures (not external political accountability mechanisms).

3.3.1.2 RBA effects on policy processes

Completion has been a focus of the GOR for some time, as evidenced by both ESSP 2010-2015 and ESSP 2013-14-2017-18. During the negotiation of the MOU on RBA, the existing priorities of the GOR were taken into account. This is apparent from unpublished documents charting the negotiation process and in the comments of national level KIs who were involved (Box 15):

Box 15: Comments of national level KIs regarding the indicator of completion

"*it was already in the policy, they had already set the policy, they were setting it up already, they already knew the completion was going to improve before they signed on – so it was in line with policy."* **[KII, International Community (2015)]**

"In designing RBA we started with the ESSP indicators" **[Former Permanent Secretary, MINEDUC (2015) (for the year two evaluation)]**

Interviews across the evaluation (years one, two and three) highlighted that central government used a combination of trips to districts and town halls to communicate their priorities to districts, sectors and

[34] Towards the end of the pilot period, interviewees suggested that MINECOFIN started to hold MINEDUC to account for the shortfall in the funds – they budget for the accrual of 100% of donor funds, and hold their ministries to account if they fail to achieve that target. This is based on a misunderstanding of RBA as a modality, as suggested in year two. The suggestion is that clarity was only achieved towards the end of the process, and had no influence on the process within the lifetime of the MOU.

[35] The Government of Rwanda understand completion as defined in the EMIS document, which is the percentage of those enrolled in the previous year who re-enrol in the following year. This relies on the information collected annually by MINEDUC officials through forms sent to each school in the country and collated through the districts. The information on those sitting the national exams at the end of the cycle forms a different data set, managed not by MINEDUC but by REB.

schools and to emphasise the importance of keeping children in school.[36] The key question is how far this may have been influenced by RBA?

Where, in relation to English, RBA was used to coordinate the actions of development partners across the sector (external coordination), in the case of completion this effect was weighted internally; providing REB and MINEDUC an argument to generate support within the wider government system. The importance of this may be emphasised in reference to weaknesses of the link between the centre and the districts and management control systems, a weakness which could be overcome – as observed in the report from year two – through intensive management communication and the use of the Imihigo systems:

> *"RBA had been accepted by MINEDUC and REB, who had the responsibility to sensitise the schools and districts about its importance. This meant that future planning would have to take into account, the RBA."* **[KII, REB (2015)]**

Thus RBA was used by senior education sector figures in communication visits to the districts, to emphasise the importance of completion since it would bring the government funds. The suggestion from participants in year three of the evaluation was of some limited awareness at the district level of the RBA, in relation to completion. This was noted in consultations with two district level officials in 2015. This marks a minor change from previous years of the evaluation where very limited knowledge of RBA outside of the highest levels of MINEDUC/ REB was identified. It suggests that pecuniary interest might have generated some minor incentive effects.

In the context of the above, it is concluded that RBA had a **weak-moderate effect** on the drive at the district level for improved completion – not through RBA as an argument in itself, but as a contributing incentivising factor in the communication efforts from the centre which sought to focus the attention at district and sector-level on preventing drop-out. This was the third of the drivers for improved completion at S3 (behind a cohort effect as a result of lagged take up of 9YBE and the 2013 REB directive on the need for a completion certificate to ensure progression). This driver may have worked to add a very slight incentive effect, impacting a small number of students.

Table 9: Contribution of RBA

Factors impacting *above trend* completion (S3 2013 and S6 2014)	Contribution of RBA to policy processes	Contribution of RBA as an incentive on the leadership of the education sector
S3 2013: Lagged effect of 9YBE	**None:** drivers already in place and independent of RBA.	
S3 2013: REB directive on completion certificates	**None:** This type of 'sitter' was not the kind intended by the RBA agreement. There is no evidence of 'gaming' to attract RBA funds, but neither can this be excluded.	
S3 2013: Cross-community efforts to reduce drop-out (increase completion)	Communication process to districts **moderately** effective and **weakly/moderately** incentivised by RBA. RBA's effect as an argument within the communication process **– very weak** – very few at district level were aware of the RBA as an argument.	Pecuniary incentive: Communication efforts **weakly/ moderately incentivised** by RBA, based, in part, on a misunderstanding of the RBA payment mechanism.
S6 2014: Cohort effect due to UPE, 9YBE, 12YBE	**None:** drivers already in place and independent of RBA.	

[36] As discussed in section 3.3.1 these town hall meetings were also used to stress the importance of language proficiency.

Alignment with Perakis and Savedoff, 2014	**Discretion**: No evidence – limited time to finesse policy, and already locked-in.	**Pecuniary interest:** Some evidence – weak incentive effect.
	Attention: No evidence – Data not new and poorly understood.	**Accountability:** No evidence of use of completion.

3.3.2. English language proficiency

FINDING 10	RBA had a strong effect to increase the urgency of GOR policy-making, which moderately reinforced the sector leaderships' efforts to stabilise the focus on English proficiency. It had a mixed contribution to their coordination of international actors (depending on the actor), and very weakly reinforced their communication efforts at the district and school level.

As discussed previously, the independent verification team recommended a payment of £1,178,100 GBP be paid to GOR on the basis of improvements in English language proficiency in the period to 2014. Changes to the test methodology have made it impossible to judge whether, or to what extent, there has been any real change in teachers' proficiency. Nevertheless, in line with evaluation objectives around 'learning lessons' it is considered important to explore the mechanisms through which change was pursued (see section 3.2.3 previously) and the influence of RBA in this the domain.

Section 3.2.3 identified four areas of intervention which sought to improve language proficiency over the course of the RBA pilot. Any possible contribution of RBA on teachers' proficiency in English must be understood through its influence on these interventions, since it is these that should be responsible for any change in proficiency (observed or perceived).

The following discussion explores the nature and contribution of RBA to any possible increase (as noted above, no scale is possible). It does so by addressing the top-line categories identified in Box 14 first, in relation to the operation of incentives, and second on RBA's influence in policy processes. As mentioned previously (section 3.2.3), it was REB who insisted upon the inclusion of teachers' proficiency in English as an RBA indicator. REB officials, consulted as part of this evaluation, indicated that they made decisions on RBA indicators according to two criteria:

> "*We were concerned with two things when we were talking with DFID: (1) we were concerned with indicators that are likely to change in a positive way – ones that you think will go up, so that you can get the money; (2) but we also wanted ones that were helping to inform you also. So it was both*." **[KII, Senior REB official (2015)]**

However, in relation to English language, it was by no means clear that the former would transpire:

> *"[The senior government officials] had gotten the feeling that they were doing poorly – actually, some of them […] saw no need for big survey, they knew that they weren't doing well."* **[KII, International Community (2015)]**

One international observer described the decision to include English as "*a brave move"* given the perceived low start at the time, while a government official described it as "*shooting ourselves in the foot".* There was, moreover, considerable resistance from DFID to the inclusion of this indicator – their aim was to include a limited set of indicators (Drew 2012, paragraph 13). Yet, English language proficiency was included in the MOU at the insistence of the negotiating team of the Government of Rwanda, and a key senior REB official in particular.

The following sections outline why this indicator was selected. In doing so, they suggest a move away from viewing RBA purely as a means of DFID incentivising the GOR, towards a view of RBA as perceived by the GOR's education sector leadership. RBA can therefore be viewed as an instrument through which the leadership of the GOR education sector sought to reinforce their aims within the government structure *using* the RBA MOU as an argument within the sector. This is a somewhat

different emphasis to the one envisaged by the prevailing theories of change for PbR/RBA.[37] It is through this lens that both the incentivisation effect (3.3.2.1.) and the way that incentivisation effect was harnessed in the policy context (3.3.2.2.) may be understood.

3.3.2.1 Creating incentive effects

This section considers the way in which RBA acted as an incentive on the leadership of the GOR. In relation to the RBA pilot in Rwanda, there are three ways in which RBA is identified as incentivising the senior leadership of REB and MINEDUC. Each of these are interlinked, and difficult if not impossible to tease apart from each other. Each, moreover, works to reinforce existing government incentives and priorities – and should therefore be seen as serving to reinforce the leadership's use of the mechanism, as outlined in section 3.3.2.2.

RBA and pecuniary interest

The functioning of RBA in relation to pecuniary interest is acknowledged in the previously cited theories of change for PbR (see section 1.2.1 (Perakis and Savedoff, 2014)), While this is evident to some extent in relation to the pilot in Rwandan education, it is difficult to tease out the operation of any pecuniary incentive over the other forms of incentive, namely that of encouragement within the government hierarchy. Certainly, in year three of the evaluation some officials suggested that the money itself was an incentive for them (Box 16).

Box 16: Comments on the monetary incentive

"*Initially, there was a sense [RBA] gave us some encouragement and excitement – it was better to [achieve the goals], since that would give us additional funds."* **[KII, Involved official (2015)]**

"the additional incentive was useful – it's natural, you always need an incentive. You either use an intrinsic motivation, like it is good to be doing well at these things; or you need external rewards. At our level, with such a huge list of priorities, this external reward was useful". **[KII, former DG REB (2015)]**

In spite of the views of some high-level officials, other evaluation evidence would suggest that pecuniary interest in relation to the RBA pilot in Rwanda, was weak at most. This is demonstrated by the fact that the evaluation has consistently detected a misunderstanding about how the funds would be channelled. While earmarked for the education sector, RBA funds were to become part of the discussions between MINEDUC and MINECOFIN and therefore might not necessarily offer additionality **[KII, MINECOFIN Official (2015)]**.[38] Many were unaware of this, and unaware that MINEDUC would almost certainly be facing a shortfall since MINECOFIN refused to top-up in-year shortages – although during the course of this final year, the point became clear. Thus one senior REB interviewee observed:

> *"But now, we saw that RBA, yes, some achievement has been made, and more money is given, and I was thinking that the money can come to that specific activity – that the money can be used specifically for English language and drop-out; but the money has been given to the central budget; it can be used for other activities – I do not see the channel to English literature and completion – I didn't see that correlation. But in a general sense, the Government of Rwanda is excited to get the money. We do everything we can: and is that not also an impact?"* **[KII, REB Official (2015)]**

In this final year of the evaluation it has become clear that there was a persistent misunderstanding about the manner by which RBA funds were channelled; this confirms findings in year two. Partly this

[37] It suggests moving the emphasis away from a simple form of incentivization where the principal (DFID) incentivizes the agent (GoR) through the promise of funding disbursal. This downplays the relative influence of the parties in defining the targets, and the distribution of power across it. The situation proposed in this article is one where the agent (GoR/REB) persuades the principal DFID to disburse against targets that will allow GoR/REB to influence other actors within the wider sector in a way that supports the GoR/REB agenda – and in the case of English proficiency, in a way that was only reluctantly accepted by DFID.

[38] That is, it was part of the budget envelope agreed by MINECOFIN with MINEDUC. The lack of additionality is because if RBA were not there, the shortfall might be made up by taxpayer funds – depending on MINECOFIN's appraisal of the MINEDUC priorities within the wider government budget negotiations.

is because the RBA agreement was explicitly understood to be experimental, and the senior leadership felt that its effects would evolve **[KIIs, REB officials, (2015)].** When senior officials in REB – the key implementers – were asked about the way the funds were channelled, they directed the evaluation team to MINEDUC. Any psychological impetus from the financial incentive was, in practice, vague and may have been lessened considerably had its true nature been clear from the outset.[39]

RBA and increased urgency from the baseline

This section concerns the effects of the baseline assessment of language proficiency which was undertaken to inform the pilot and form the baseline against which improvements could be measured. As reported in previous years, there is a perception that introducing processes of verified and validated externally-produced data acted as 'wake-up' call for GOR, driving the commitment of senior education sector leaders to strive for progress. Comments from across the evaluation period include:

Box 17: Comments on the baseline

"RBA is seen as a learning process to assist us to achieve our own aims ...it provided a fresh momentum...like a bell ringing". **[KII, REB official (2014)]**

[…] if there hadn't been an RBA agreement, and if you hadn't had the baseline, then it would have been different – no-one could ignore [the baseline], and that did promote the indicator. It got on the agenda." **[KII, International Community (2015)]**

"I have never seen the government so focused on results. It [RBA] has definitely been a thing that has focused attention. **[KII, International Community (2015)]**

"If it is something that [the GOR] wanted to do, then they'll get it done – they don't like negative results, they will address the issues if these are forthcoming" **[KII, International Community (2015)]**

As discussed, the findings of the baseline assessment of language proficiency were poor. Government officials, when interviewed in 2015, stated that they had expected the results of the baseline, that it had not been a surprise. Comments from the international community – including those who were in the room when the detailed results were announced – suggested that to the contrary there was surprise at just how low the results were. At the time of the RBA agreement, evidence on the competence of teachers in English was very weak. As one member of the international community suggested:

> "*They knew they had a problem, but they had no data on the scale of the problem that they could point to*." **[KII, International Community (2015)]**

The baseline prepared by the British Council revealed a worse situation than originally presumed. The baseline therefore acted to focus attention on the need to improve teacher proficiency in English.

> *"For us the baseline was not evaluative but diagnostic – that is very important. That was what we said in the TV and to teachers, that it was not evaluative. It was to establish for us the level of the groups..."* **[KII, former DG REB (2015)]**

While there is no further evidence to suggest that the study was used for a detailed restructuring of the SBMP (within the timescale of the MOU), **within the context of a locked-in policy, RBA had a strong effect in increasing the urgency of the target, intensifying the incentivising effects.**

RBA results as a management incentive

The effect of RBA in driving accountability of Government (through making results visible) is central in arguments supporting RBA (see Box 17). In the case of the RBA pilot in Rwanda, as with completion (discussed subsequently), the challenge with RBA as a means of holding to account is linked to teasing it apart from other incentives and control mechanisms. The year two evaluation report devoted

[39] With a contracting budget, it may be that the size of the tariffs would have been sufficient to drive an incentive to MINEDUC and REB, had their contribution to the budget been adequately understood. However, given the way in which RBA funds were disbursed, REB officials have suggested that the incentive of the tariffs is greatly lessened.

considerable space to education sector management controls in Rwanda – noting stronger controls in relation to completion than English. Nevertheless, knowledge that there would be an endline test intimated the production of a rigorous data set which would not be manipulated and therefore would allow the senior leadership of the sector to be held to account within the GOR hierarchy. This appears to have had a moderate to strong effect on the psychology, and served to sharpen the focus.

3.3.2.2 Effects on policy processes

RBA is now seen as having an incentive effect on the leadership, particularly through the urgency generated from the baseline. This section explores the effect that RBA had on policy. To understand this, it is important to understand that the time difference between the baseline and the endline was only two years, giving limited time for policy innovation. Moreover, the key choices had already been made: the government was locked into EMI and into the School Based Mentoring Programme as the key means of delivering upon it. The strategy for this programme emerged from a conference in 2012, before the baseline survey of language proficiency. The baseline simply established a diagnosis of the scale of the problem; there was no evidence to suggest that the study was used in any real way to restructure the SBMP – rather it acted as a trigger to unite the sector behind the programme and to pursue the existing policies with greater urgency.[40] The nature of the 'attention' and the 'discretion' theories of change (Box 17) should be seen through this reframing.

RBA reinforcing, locking-in and stabilising policy

The REB leadership *anticipated* and *harnessed* the incentivising effects of RBA to further their policy goals – among which English, at that time, was crucial. As noted above, this is a somewhat different emphasis to the one envisaged by the prevailing theories of change for PbR/RBA.

In interviews with senior national level stakeholders in year three of the evaluation it was apparent that RBA supported efforts of senior REB officials to manage the constantly changing policy context and moving priorities. As one member of the international community observed:

> *"[REB] are doing a lot, but the priorities are changing […] there is a priority on drop-out; but everything becomes a priority – there's a massive, massive amount of work to do. There is a demand that things should be seen as a success – so they say, if that's a success, then why do we need to do more? The priorities change, and they move on to the next thing…"* **[KII, International Community (2015)]**

As stated, English language proficiency was a priority for the education sector over the course of the RBA pilot. Senior officials wanted to *use* RBA as an incentive for the sector as a whole (see below) and to sustain the focus over time. Specifically, senior REB officials harnessed the RBA incentives to stabilise the focus of the sector on language proficiency, and ensure that, for the duration of the RBA agreement, the education sector would be focused on this. RBA lent senior officials the argument to reinforce the importance of this priority, amongst others. This is illustrated by the following quote:

> *"…We used RBA to pursue English Language Proficiency and Completion goals. […] the purpose was to sustain the momentum. We conducted the town hall meetings and 'kubaza bitera kumenya' [live TV and radio phone-in programmes] in addition to the town hall meetings – to make sure what we have started continues."* **[KII, former DG REB (2015)]**

However, already priorities were shifting as the MOU drew to a close. It is important to note that this process of locking in the sector has also had some negative implications: thus the relatively positive results from the endline risk suggesting that the 'job has been done'; furthermore, by locking in, there are risks that other beneficial targets, such as pedagogy, were excluded (see section 3.3.3.).

RBA contributing to sectoral coordination

Evaluation evidence also suggests that senior education officials used the incentive of more funds, and the accountability of an independently verified endline assessment to provide a psychological incentive

[40] This was not the purpose of the baseline. Subsequent diagnostic tools designed by a volunteer were used for this purpose.

to the wider government apparatus, contributing to greater coordination in policy and programming across the sector (from the centre, down to district, sector and school-level). The link between the central government and the local government apparatus – including the crucial link between REB and the (then-titled) District Education Officers – was in general weak, with a lack of district-level planning in the education sector in the manner envisaged by the ESSP (CfBT 2015, 16). However, as indicated in the year two evaluation, where REB and MINEDUC officials launched specific communication efforts, this weak link could be overcome.

It is suggested that REB used RBA as an argument and an incentive to reinforce and push the focus on English during visits to the districts by REB/ MINEDUC officials in 2012 (see also Upper Quartile, 2014).The incentive is therefore used as a policy effect. As with previous indications, however, the year three evaluation suggests very few officials at the district or school level knew about RBA.

Box 18: Comments on the policy process

"In 2009 we moved from French to English. We knew it would take a long time to make the shift. The skills of teachers needed to be improved. The RBA seemed to be an incentive. There is always resistance to change but at the policy level it seemed a good idea. We used it to sensitise teachers, Head Teachers, Vice Mayors and DEOs to the priorities. To do this we were moving around the provinces and conducting town hall meetings, we met various different kinds of groups of people and we informed them about the importance of RBA. We went round the schools […]." **[KII, former DG REB (2015)]**

A further key point is that it was not just the GOR that was incentivised to focus on English language proficiency; by persuading DFID to put this into the MOU, GOR created an incentive not just for themselves but also for DFID and ultimately the wider international community.

A DFID-Rwanda education official (consulted in 2015) suggested that they had not originally intended to be involved in the level of policy detail with regards to the question of English language proficiency that they ultimately were but had done so "*because of RBA, and because REB saw DFID as a key actor*".[41] DFID officials were important resources within the policy environment. They chaired the Sector Working Group and the Technical Working Group that focused on Teacher Development and Management, including their English proficiency. DFID were also able to bring in consultants and funding for key coordination and research functions – including inputs to discussions around a wider and more holistic approach to EMI (resulting in development of a costed strategy that was ultimately unable to secure funding and was not implemented).[42]

Further, the key conference in Gisane in 2012, which refined the SBMP, brought together USAID and other NGOs to build support in this area of programming. Interviews with representatives from the international community (undertaken in both years two and three of the evaluation) suggested that, where possible given the priorities of their own organisations, the international actors came on board in a more coordinated fashion **[KIIs, International Community (2015)].** The degree to which their own priorities permitted international actors to adopt the REB and MINEDUC priorities is therefore different, and the effect of RBA on this therefore mixed. One KI from the international community commented:

> *There have been presentations at every JRES since RBA came in. [ ] It has certainly driven programmes, it has driven SBMP. Had it not happened, along with L3, never been as much of a [drive] around it to try to action it."* **[KII, International Community (2015)]**

It is useful to see RBA as an instrument that could be used by the leadership to push for implementation of their existing programme, rather than as an opportunity for policy innovation. It should be noted that

[41] The DFID official also noted that the decision to become involved had been the correct one, since the issue of EMI was crucial to teaching outcomes within Rwanda.

[42] National strategy for Teachers' In-Service Training to support English as the Medium of Instruction (EMI).

growing disenchantment with the mentoring programme resulted in its cancellation shortly after the RBA MOU expired.

3.3.2.3 The contribution of RBA to English language proficiency – in summary

Table 10 summarises the contribution of RBA in relation to English language proficiency. It highlights the four areas to which the evaluation's analysis has sought to identify a contribution and the nature of any observed contribution.

Table 10: Contribution of RBA

<table>
<tr><th>Factors impacting teachers' proficiency in English</th><th>Contribution of RBA to policy processes</th><th>Contribution of RBA as an incentive on the leadership of the education sector</th></tr>
<tr><td>School Based Mentor Programme</td><td rowspan="3">Contributed to REB senior leadership's efforts to
Stabilise and focus policy effort at national level (moderate)
Increase sectoral coordination through communication efforts to the international community (mixed)
Increase sectoral coordination through communication efforts to district officials (very weak)</td><td rowspan="3">RBA seen to reinforce incentives of the education sector leadership by:
Providing (misunderstood) fiduciary incentive (moderate).
Focusing attention on the state of language proficiency, providing an important wake-up call for, and urgency in, the leadership (strong).
Reinforcement to the existing management incentive through credible endline measurement (strong).</td></tr>
<tr><td>Teacher training</td></tr>
<tr><td>Teacher focus on English</td></tr>
<tr><td>L3, VSO, other NGO support</td><td>Contributed to REB efforts to coordinate development partners – DFID, USAID and NGOs – behind efforts in the area of English language proficiency (mixed).</td></tr>
<tr><td rowspan="2">Alignment with Perakis and Savedoff, 2014</td><td>Discretion: No evidence, very limited time/ space for innovation.</td><td>Pecuniary: Weak evidence of effect</td></tr>
<tr><td>Attention: Some evidence as a result of baseline but in terms of incentivising focus on existing policy.</td><td>Accountability: No evidence</td></tr>
</table>

3.3.3. Unanticipated or unintended effects

FINDING 11	The endline for English proficiency may have been interpreted wrongly to signal *'job done',* which could cause considerable harm since evidence suggests much still needs to be done.

FINDING 12	RBA created unanticipated effects; in particular through locking-in the government and thereby legitimising certain programme decisions which may exclude other policy priorities.

Three main unanticipated effects can be observed from the foregoing discussion.

Planning effects of RBA: The first is that RBA leads to fluctuating revenue, which may affect planning. As in the year two report, MINECOFIN confirmed that in budget discussions with MINEDUC, they calculate RBA the same way as they calculate the EC variable tranche; that is that they plan for the disbursal of 100% of the funds. This is instituted in their formulation of MINEDUC's budget ceiling, and

thus the annual budget for the education sector overall.[43] Given the likelihood that the GOR was not going to achieve the maximum RBA pay out, this means that the sector would face a shortfall in-year to that extent. MINECOFIN made clear that they would not make up any such short-fall in-year: the sector would have to bear the reductions and either find other funds or cut their activities. RBA therefore created an inevitable uncertainty in funding, which must be borne by the sector. The fact of this shortfall is mitigated by the degree of fluctuation in the education budget as a whole. In an interview with a senior MINEDUC official involved in the planning and budgeting process, the degree of fluctuation of activities overall was emphasised – the shortfall would be accommodated in that general fluctuation.

Signalling with non-comparable data: The second unanticipated effect arises from the erroneous lesson taken from the endline results on English language proficiency. While RBA locked-in the focus on English proficiency in the context of a fluctuating policy environment, this message was being interpreted as signalling 'job done'. This in turn risks signalling that it is appropriate to 'move on' i.e. to consider the priority achieved, and to move focus to one of the other of the range of pressing priorities within the education sector. This risk was stressed in particular by members of the international community, who fear that the endline has allowed the new leadership of REB to claim that they have achieved their goal. The risk of this message was made possible by two aspects of the way the endline was applied:

1. By the drop in the standard for payout to B1, which is significantly below that of the level of competence needed to teach in English; and
2. By the nature of the baseline test, which was considered to be unfair in its application to Rwandan teachers.[44]

Interviewees from the international community perceive that the state of English proficiency is far below an adequate level for teaching purposes; there is a great deal of work still to be done. Insofar as the endline has allowed communication of a different message, this was a negative effect of RBA.

Box 19: Comments from the international community

"[One key REB official] responding to the endline results 'oh, it's fine, we don't need to do anything more'" **[KII, international community, (2015)]**

"The priorities change, and they move on to the next thing e.g. the Aptis Endline, we were fearful that it was seen as 'job done' – the publication was treated as a celebration of success – what you should do is set another indicator, more challenging […] Everyone is so busy, and no-one coordinates across the piece. There are some concerns that the [leadership] thinks it's job done." **[KII, international community (2015)]**

Negative effects of locking-in to a questionable programme: As observed above, RBA contributed to locking-in the government to focus on the RBA indicators. However, interviewees noted that this had both negative as well as positive implications. It meant that they locked-in to the exclusion of other education priorities. There were some strong concerns voiced by members of the international community that the focus on English competency which the RBA helped to lock-in acted to exclude a wider focus on pedagogy (this was contested by GoR interviewees). The suggestion is therefore that RBA to some extent legitimised its continued focus on the particular approaches that had been adopted – in particular a concern that it legitimised the School Based Mentoring Programme, which was contested in policy discussions within the relevant Technical Working Groups.[45] The subsequent

[43] It should be noted that all donor aid provisions must be made to MINECOFIN, and are therefore part of the overall negotiation between MINEDUC and MINECOFIN – even if they are earmarked to the sector or a particular activity, there is no necessary additionality, since MINECOFIN can simply withdraw tax revenue to the sector as a whole.

[44] It should be noted that both of these were considered at the time, but the risks of very low payout and an unfair test were considered to be greater.

[45] REB has just made the decision to reform the SBMP, triggered in part by a change in leadership. The justifications were the excessive cost of SBMP in the context of a budget cut and the lack of effectiveness of the programme.

substantial 'restructure' (in practice an almost complete redesign) of the SBMP seems to confirm this. This issue is a product in part of the stage in the policy cycle at which RBA was agreed, and in part the short length of the MOU's term, which between baseline and endline spanned only two years. Combined, it meant the space for policy innovation was very limited, and became more a matter of the evaluation of the existing policy commitments.

3.4. Value for Money

3.4.1. What is Value for Money in relation to RBA?

A core question for the evaluation of RBA is whether it represents good value for money (VfM). VfM is principally used as an ex-post term, put simply 'What did we get for our money?' and 'How does this compare to what we could have got elsewhere?' aiming to inform the follow-up: 'Was it worth it?'

For the RBA pilot this means whether the benefits of applying aid through the mechanism of RBA outweigh any additional costs. An immediate complicating factor is that the benefits of education accrue in the future, so assumptions must be made on how and where these benefits will form, while most of the cost is incurred in the present. This necessitates modelling of education returns.

In establishing the VfM of RBA against a counterfactual of providing aid to the Rwandan education sector through another means – such as sector budget support (SBS) – RBA is required to generate benefits that other means do not provide.

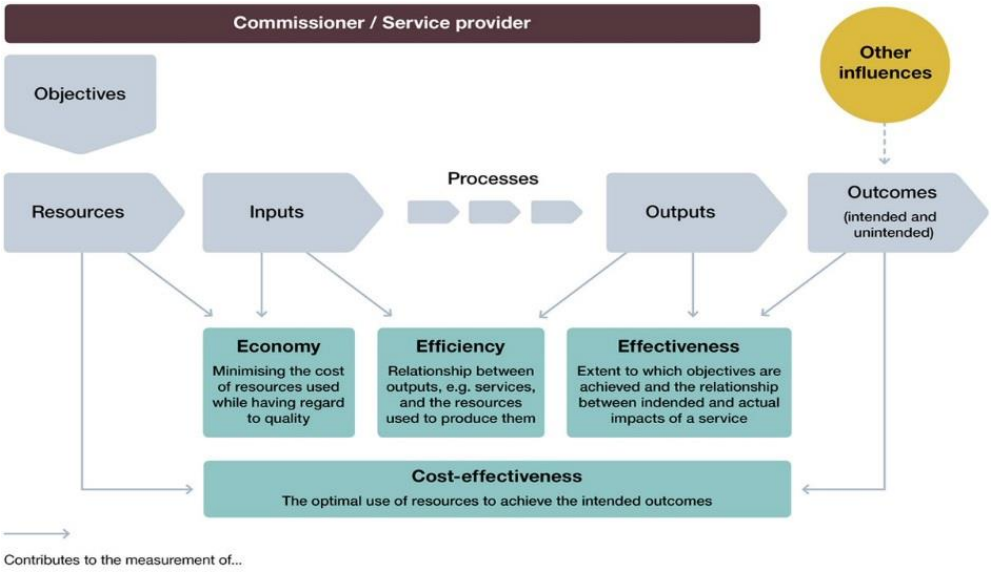
The estimated benefits of RBA should be founded on evidence of it providing an effect leading to increasing levels of completion at key stages and overall level of education achieved. This may be evident through, for example, improving or strengthening GOR policy implementation consistent with greater quality and quantity of education, or improved quality directly through incentives to improve the level of teachers' English proficiency (therefore enhancing the classroom experience in this language of instruction).

The costs of RBA as opposed to another modality of aid to the education sector include the costs of this evaluation itself, as well as the costs of independent verification of results (the trigger for disbursal of payments). In addition, there may also be costs associated with the modality of aid itself. In particular that the RBA mechanism, as designed for this pilot, may provide uncertainty in terms how much money will come to MINEDUC according to performance. There may be costs to this uncertainty if it impedes the Ministry from planning effectively for upcoming education expenditure (see 3.3.3 above).

3.4.2. An approach to assessing VfM

The VfM equation is usually split between the '3-Es' as shown in Figure 4 with sub-questions on *economy*, getting a good price for inputs; *efficiency*, getting sufficient outputs for activities undertaken; and *effectiveness*, achieving overall goals and objectives. A fourth criterion combines the approaches as *cost-effectiveness*, i.e. outcomes achieved are commensurate with spending.

Figure 4: UK National Audit Office (NAO) approach to VfM



Source: NAO website (accessed August 2015)

The framework underlying the standard '3-E approach' is the linear theory of change set out by the arrows from objectives through to outcomes in Figure 4. The idea is that with the exception of 'other influences', resources are under the control of a service provider or agency and can be clearly traced along the chain. In designing the VfM assessment for the RBA pilot in Rwanda, it was concluded that this linear approach is not possible for evaluating a payment by results (PbR) mechanism. This is because payments are intended to incentivise (in some way) performance by the partner agency, in this case institutions of the GOR, but through no pre-specified channel. The theory of change is therefore open in that inputs or even outputs cannot be clearly specified either before or after the pilot, so the '3-E approach' is not easily applied.

As set out in detail in Appendix 9 on VfM, an approach was developed for the assessment of the VfM of RBA to education in Rwanda, with a model developed for the year two evaluation report in 2014. The approach has been continued for this final evaluation report; with some adaptations and extensions to try to ensure greater realism and rigour of the estimates produced. For this final report, the model has also been run for all the three years of the RBA pilot – 2012, 2013 and 2014.

As set out in the year two evaluation report and annex (see Upper Quartile, 2015), this assessment does not follow the 3-Es approach for the reasons cited above, but instead looks at the cost effectiveness of RBA relative to not providing RBA. There are two components to this, based on two different counterfactuals:

a) The VfM of aid spent on education through the RBA instrument, compared to the counterfactual of not providing that aid to education.

b) The VfM of aid spent on RBA, compared to the counterfactual of still providing that aid to education, but by another instrument without any explicit incentive present – for example via sector budget support (SBS).

The overall VfM of RBA is then A+B: that is, the effectiveness relative to the cost of the aid itself (A); and the effectiveness of RBA relative to providing the aid but in another form (B). The test of most interest in looking at RBA as opposed to other aid modalities is B, as this can be taken to be the narrower test of the hypothesis that RBA functions through an incentive effect that is not present in other forms of aid. The model presents two versions of test B:

- The first, **B1**, is a 'naïve' estimate using the benefits derived from the number of 'extra sitters' at P6, S3, and S6, which the RBA design pays out upon, i.e. assuming a zero trend, which is not at all likely to be there in reality.

- The second, **B2**, uses the econometric model of the RBA evaluation to estimate the number of 'additional sitters' at these grades – i.e. the *statistically significant* level of change, taking into account changes above or below existing trends – and is therefore the more rigorous of the tests.

3.4.2.1 Details of the benefits from the VfM model

Measuring the benefits to education represents a complex problem. The model used takes the principal benefits of education to be in the form of economic 'returns to education', following the economics literature. In order to model these benefits, a model of the Rwandan education system and labour market was built; this was necessary to measure benefits that accrue when the educated are in employment, and then for many years afterwards. The model included the following components:

- **A model of the existing cohort within education.** In a country such as Rwanda, many different ages are represented at each grade, particularly in primary school. Modelling the ages of the current cohort in the education system was important to understand how long they would be in the labour market, assuming work starting at age 16 for those not in schooling, and continuing to the national retirement age of 60 years old. Mortality rates were also added to the model to ensure greater realism.
- **Estimates of how long each individual stays in education.** The model uses existing estimates of repetition, progression and drop-out rates at each stage of education, varying these rates for sensitivity analysis. The model shows a probability distribution for when each member of the cohort will leave education and enter the labour market, and crucially how many years of education they will have when they become economically active.
- **A model of the labour market, and returns to education in the form of wages.** Once an individual enters the labour market, their wages are estimated based on the year they enter and how many years of education they have. This uses estimates of returns to different years of schooling, assumptions on economic inactivity rates, and real wage growth across years.

The model is then used to estimate the benefits of education from both test A and test B. In the case of test A the approach is to look at the overall benefits to education for the given year in question, with an assumption that the aid disbursed contributed a proportionate share of the benefits. For test B, the model is run to estimate the education outcomes and then returns to education for individuals that successfully complete at grades P6, S3 and S6, and who are assumed to not have continued in education in the counterfactual.

3.4.2.2 Details of the costs from the VfM model

The costs modelled by the VfM model are somewhat simpler than the benefits calculations. For test A the cost is assumed to be the amount of aid disbursed. This is compared to the overall costs to education for the given year. The proportion of the costs funded by the RBA disbursement is used to estimate the proportion of the overall benefits to education for the year in question by the aid disbursed.

For test B, costs are those that accrue due to using the RBA mechanism as opposed to disbursing the aid by another means. This includes the evaluation costs and verification costs of the RBA pilot; costs directly accrued by DFID. It also includes costs of education for those individuals, who would otherwise have dropped out, who continue in education. These costs accrue in future years and are assumed to come from the GoR, from aid to education, and from the portion of education expenditure coming from private means and households.

3.4.2.3 Assumptions in the VfM model

Despite attempting to model based from real data, a model such as this projecting so far into the future is inevitably built on a large number of assumptions. These aim to be as justified as possible, but they should be clearly stated and noted. These include:

- A discount rate of 10 per cent – this is the standard DFID discount rate, including for Rwanda. As the model seeks to measure returns to education in the labour market from those in education now, the returns are often far into the future, while most of the costs are incurred in the present, benefits are therefore more heavily discounted than costs.

- Drop-out rates, repetition and progression rates all continue into the future. This does not take any account of changes to supply, quality or demand for education in coming years.
- There is no drop in quality that would impinge on the likely returns to education of individuals leaving schooling at various levels. The underlying assumption of relatively constant educational quality was not firmly investigated given the difficulty of doing so. This is also outside of the remit of the current evaluation.
- Labour demand will continue to be sufficient to absorb increasing amounts of individuals in the market with an average higher level of education, with no significant changes to unemployment, economic inactivity or underemployment.
- Much of the assumed benefit in the VfM model is derived from the implied extra years of schooling over and above the incentivised year. If extra schooling for additional sitters was found to be lower than for other students – i.e. an 'additional sitter' at P6, S3 or S6 was subsequently more likely to drop-out than other students – then the size of the benefits would fall substantially.

The essence of the assumptions made in the VfM model is that current trends will continue into the future. In reality, change can be discontinuous. With respect to RBA, it is very important to note the central assumption that RBA itself does not represent a discontinuity – i.e. RBA does not change behaviour around other aspects of the system that would impact on educational outcomes.

3.4.3. RBA in Rwandan Education – Value for Money?

FINDING 13	The VfM assessment found that, in the absence of any negatives, only a small number of additional completers would be needed to justify the small additional costs of RBA.

FINDING 14	In the case of the RBA pilot in Rwanda, the findings generated by the VfM model should be treated with caution as the wider quantitative and qualitative evidence base casts doubt on some of the models' assumptions. Specifically, the holistic evidence base suggests that cohort effects and wider GOR policy implementation were the main drivers of increased completion – as such the increase is not wholly additional. This means the benefits from Model B2 could be much smaller as changes cannot be attributed to RBA.

FINDING 15	There is some evidence that the GOR incurs a cost as a result of the uncertainty of the RBA modality – i.e. how much funds will be allocated and therefore planning expenditure profiles for coming years. However, the level of this uncertainty cost, as well as any opportunity cost of a funding short-fall to MINEDUC in comparison to the greater stability of SBS has not been possible to model.

The VfM model estimates the cost effectiveness of aid to education via RBA based on the net present value (NPV) of discounted benefits and costs calculated. The model is run for each of the three years of the RBA evaluation, and the overall results of this are shown in Table 11.

The overall finding is that aid to education in general provides good VfM (results for test A), with a benefit-cost ratio (BCR) of 2.6, and an estimated £11.2 million in benefits accruing as returns to education from £4.4 million disbursed in aid giving a NPV for test A of £6.8 million. However, to emphasise again, this is against the counterfactual of not providing aid, so the RBA mechanism itself is not what is responsible for this net benefit, but rather the aid itself.

In terms of the more important test of RBA as an instrument, the naïve estimate (test B1) finds significant and large benefits, with a BCR of 2.2 and net present value (NPV) of over £100 million for the project period. However, this is based on unreliable statistical assumptions such as a zero trend, and does not take into account negative changes to cohort completion rates during the pilot period.

The more plausible test, test B2 has a smaller result with a NPV of £18 million and a BCR of 1.1 – though it should be noted that this includes costs that are not borne directly by DFID: excluding these costs would make the BCR much larger. The narrow result for B2 is in part due to statistically significant negative changes added to the model (as requested by the evaluation reference group), with 2012 generating a negative NPV, only partially offset by the changes in 2013 and 2014. The key result here though is still that there is a significant benefit from the application of the RBA mechanism, based on the quantitative findings.

Combining the two models gives an overall assessment of the VfM of RBA, with a BCR of 2.2 for A+B1 and 1.2 for A+B2, and a NPV of £117 million and £25 million for the two combinations respectively. Sensitivity analysis is conducted in Appendix 9 finding that NPV returns remain positive even when a number of assumptions are made more stringent.

Table 11: Summary results from VfM assessment of RBA for all years (GBP million, in 2015 GBP) [46]

Test	PV benefits (£ million)	PV costs (£ million)	NPV (£ million)	B-C ratio
Test A				
Benefits attributable to the 2012 year of education	3.0			
RBA disbursed for 2012		1.2		
Benefits attributable to the 2013 year of education	4.9			
RBA disbursed for 2013		1.9		
Benefits attributable to the 2014 year of education	3.2			
RBA disbursed for 2014		1.2		
Overall NPV for test A	**11.2**	**4.4**	**6.8**	**2.6**
Test B1				
Net benefits from extra sitters in 2012	65.6	34.7	30.8	1.9
Net benefits from extra sitters in 2013	74.5	33.4	41.2	2.2
Net benefits from extra sitters in 2014	63.2	24.5	38.7	2.6
Evaluation costs and verification costs		0.7		
Overall NPV for test B1	**203.3**	**93.4**	**110.0**	**2.2**
Test B2				
Net benefits from extra sitters in 2012	21.1	52.1	-31.0	0.4
Net benefits from extra sitters in 2013	72.7	41.0	31.7	1.8
Net benefits from extra sitters in 2014	82.0	64.1	17.9	1.3
Evaluation costs and verification costs		0.7		
Overall NPV for test B2	**175.8**	**157.9**	**17.9**	**1.1**
Tests A and B combined				

[46] The benefits and costs from 2012 and 2013 are uprated into 2015 GBP using the UK Retail Price Index (RPI).

Overall A+B1	**214.5**	**97.7**	**116.8**	**2.2**
Overall A+B2	**187.0**	**162.3**	**24.7**	**1.2**

3.4.3.1 Cost effectiveness per sitter

Given all of the assumptions noted, it is instructive to look at the individual VfM for the changes that RBA intends to bring about. Table 12 provides the unit breakdowns for test B2 in terms of how much benefit is derived from an *additional sitter* at each of the grades, P6, S3 and S6. This is informative in terms of the scale of change that is required to make RBA good VfM. It shows that the NPV per individual additional sitter ranges from around £1,000 to £2,500, with this benefit the additional lifetime earnings as a result of extra years of schooling, offset by the cost of that schooling. This compares well to the tariff of £50 or £100 depending on the grade paid out by the RBA mechanism.

The result of this is that a relatively low number of additional sitters are required to break-even on the additional costs from the RBA mechanism – the evaluation and verification costs. Thus, just under 500 additional individuals completing P6 would be the break-even point for the whole programme, compared to just over 200 additional individuals completing at either S3 or S6. This would be an average 0.3 per cent additional completion rate for sitters in either of these three grades over and above the counterfactual to break-even. This shows that if RBA is shown to have a significant and positive incentive effect, then even a small-scale impact on completion could be VfM.

Table 12: Analysis per additional sitter for test B2

Additional sitters at:	PV benefits	PV costs	NPV per individual additional sitter	Number of sitters required to break-even on additional RBA costs
P6	£2,976	£1,856	**£1,120**	**493**
S3	£4,499	£2,079	**£2,419**	**228**
S6	£4,080	£1,584	**£2,496**	**221**

3.4.4. RBA in Rwandan Education – Do the assumptions stack up?

The preceding argument shows that, according to the assumptions of the model, VfM is achieved even if RBA causes only a small number of extra exam sitters. However, the findings from the quantitative and qualitative strands of the evaluation cast significant doubt over some of the assumptions made for the VfM modelling. Notably:

- The cost from uncertainty of RBA disbursements cannot be modelled. There is evidence that achieving less than £3 million per year in disbursements was experienced as a short-fall to MINEDUC, partly due to misunderstanding about how RBA would function. This uncertainty effect could offset any incentive effect generated.
- The opportunity cost of the aid disbursed is not clear, as it is not clear how the unspent funds are used. In addition there may be substitution between aid disbursed and tax revenue by MINECOFIN, which would mean this aid to education is not truly additional – this would undermine the benefit modelled for test A in particular.
- The evidence that an RBA incentive effect did exist, and the scale of any associated benefits, is mixed. This means that the statistically significant additional sitters modelled by the evaluation (the basis of the VfM assessment) are unlikely wholly *additional*; the result instead of chance, wider policy implementation, cohort effects or perverse incentives. This means the benefits from B2 could be much smaller than estimated.
- There is some evidence that the additional completers modelled to be statistically significant in the quantitative model may not be like other completers in terms of the likelihood they will continue for further years in education. Specifically, a lower number of students entered S4 than we would expect. This would significantly deflate the benefits modelled.

- Completion is not an accurate proxy for the quality of education. If resources in the education sector do not keep up with intake, quality may be declining, which would significantly reduce returns to education.
- It is possible that the emphasis on P6, S3 and S6, had the perverse consequence of taking focus away from other grades, a negative result that is not modelled here.
- Finally, assumptions about the labour market depend on many external factors, and evidence suggests that greater labour supply at different levels of education could reduce wages in future if the economy cannot keep pace via labour demand. Combining so many assumptions inevitably leads to a higher degree of uncertainty around the results.

In summary, the benefit derived from the VfM model estimated to come from DFID support to education in Rwanda through the mechanism of RBA, should therefore be treated with caution. However, if an incentive effect could be created through a PbR aid mechanism, particularly if it improved educational quality, the model shows this could easily generate very significant returns.

4. Conclusions and recommendations

This final section of the evaluation report collates the findings presented in the preceding sections to address each of the macro-level evaluation questions in turn.

Evaluation question 1: What has been achieved?

In the context of the RBA indicators, the most impressive achievement has been improvement in the absolute number of exam sitters at both S3 and S6 levels, with respective increases of 11% and 49% over the period of the RBA pilot. Performance at P6 has been virtually flat, with relatively recent increases stalling during the pilot period (2011-2014) **[Finding 1].** With regards to the English proficiency of teachers, any apparent success is undermined by inconsistent data: a different test was taken in different conditions **[Finding 8].** The stated achievements saw £5.4 million GBP disbursed as RBA over the course of the pilot period: approximately £0.8million GBP for P6, £1.9million GBP for S3, £1.6million GBP for S6 and £1.2million GBP for English.[47]

Evaluation question 2: Has RBA contributed to impact in relation to the envisaged results?

The quantitative evidence is unanimous in finding that RBA had no consistent effect on completion results **[Finding 3].** Where completion performance was found to be outside of the range of what would be expected without RBA, performance was as likely to be below trend as above it **[Finding 4].** There were two instances of above-trend performance: S3 exam sitting in 2013 and S6 exam sitting in 2014 **[Finding 5].** The latter performance was unanimously linked in qualitative evidence to the historical policy changes of 9YBE and 12YBE, a view supported by the quantitative research **[Finding 2 and 6].** The picture surrounding S3 performance in 2013 is less clear. There are three main proximate causes of the S3 increase: a lagged effect of 9YBE, concerted community- and school-level efforts to address drop-out/ repetition, and a REB directive relating to the requirement to have a junior secondary (S3) exam certificate to allow progression. The first could not be caused by RBA, as it predates the agreement **[Finding 2].** For the second, RBA had a moderate contributory incentivising effect to the communication efforts, which were weakly efficacious **[Finding 7].** The third did not concern new S3 students, but instead those who had already graduated to senior secondary; they were not the intended targets of the RBA. There is no evidence to suggest that this directive can be attributed to RBA; if it was, it would be considered an instance of 'gaming' **[Finding 7].**

It is not possible to comment on the extent to which RBA may have impacted the results for the indicator of English language proficiency due to lack of comparability of data **[Finding 8].**

Evaluation question 3: What factors have impacted on the achievement of RBA results?

As noted, the qualitative evidence suggests three main proximate factors impacted on the S3 increase in 2013: a lagged effect of 9YBE, concerted community- and school-level efforts to address drop-out/

[47] Inconsistency due to rounding.

repetition and a REB directive relating to the requirement to have S3 exam certificates. Of these, the evaluation considers that RBA may have exerted influence on the second factor only. The contribution of RBA in this instance is discussed in relation to evaluation question 5.

The four influencing factors in the area of English language proficiency were identified as the School Based Mentoring Programme, improved teacher training, the general management communication effort to the teaching profession around the importance of English language proficiency and activities of other international donors and NGOs in this policy area (section 3.2.3).

Evaluation question 4: How is the RBA approach perceived in Rwandan education?

The RBA agreement was perceived within senior elements of the GoR hierarchy as an experiment. Although officials were aware that DFID would disburse on completion and English-proficiency related targets, the precise mode of payment was not clear within key elements of the main implementing body, REB, until the last payment. The realisation that DFID disbursals under RBA were not earmarked for English language proficiency or completion per se but were instead part of the negotiation with MINEDUC, was disappointing for REB officials. More broadly, RBA was not understood well, outside a small group of senior officials. Very few people at the district level were aware of RBA – of the nine district and sector level officials interviewed in year three, only two had some notion of RBA (see section 3.3.1 and 3.3.2 for detailed discussion of these points).

Evaluation question 5: How did government respond to RBA?

In the case of both completion and English proficiency, the government was already committed to a broad policy direction, and therefore had limited span of control within which they could react or innovate. The nature of the government response to RBA was therefore defined by the way that it was used as an argument by the senior leadership of the GOR education sector. It helped senior members of the leadership to lock the government into the priorities articulated (a significant effect in the light of the constantly fluctuating policy context). RBA supported the government's efforts to coordinate the sector more broadly: the funds and prospect of measurement were used as an argument in communication at the district and school level; and they were used to help ensure the focus of DFID, a key source of resources, and the wider international community **[Findings 9 and 10].** The RBA agreement should therefore be understood as helping the education sector leadership to commit to the targets and to communicate this across the national/ local education sector apparatus more broadly, by virtue of the funding and the measurement.

Question 6: Has Value for Money been achieved?

The VfM exercise conducted for this evaluation was ground-breaking in that it applied standard VfM tools to the innovative RBA instrument across two different counterfactuals. The first counterfactual on aid to Rwandan education (as opposed to no aid) showed excellent VfM.[48] For the second counterfactual – the one of most interest to the RBA evaluation – i.e. aid provided by RBA as opposed to another modality – a naïve interpretation finds that RBA offered very good VfM. In this vein it is concluded that the VfM exercise demonstrated the potential of an RBA modality to provide VfM – with a low break-even number of additional sitters, even a small-scale impact could be VfM **[Finding 13].**

In relation to the specific figures generated by the VfM assessment in the Rwandan context, the assumptions underpinning the result were found to be problematic: e.g. that RBA *caused* increases in completion and that extra exam sitters were no different from others **[Finding 14].** Both of these assumptions were found to be questionable, and undermine the VfM of RBA in this particular setting.

Further, it is not possible, in the model, to take account of the uncertainty of RBA in comparison to SBS **[Finding 15].** While RBA was perceived to be a small amount of money, it is possible that the reason it did not receive a greater response was more due to the short length of the agreement (see section 3.3.1 and 3.3.2 for detailed discussion with regards to limited time/ space for innovation) and the way in which the funding was disbursed (i.e. the relevant department did not necessarily anticipate or notice

[48] In that the benefits significantly outweigh the costs even though benefits accrue in the future and are therefore more heavily discounted than the costs incurred in the present.

an increase in their budget as a result). In terms of the disbursements, we are confident that the performance at P6 and S6 would have happened anyway (£0.8million and £1.6million) **[Findings 2 and 6].** Furthermore, £1.2 million was disbursed for English improvements that may not have occurred **[Finding 8]**. The picture around the S3 performance which lead to £1.9million being disbursed is less clear, but the results have not led to a sustained increase in students studying at higher secondary **[Findings 3 and 7].**

Question 7: What lessons have been learned to inform RBA in Rwanda and elsewhere.

In responding to evaluation question seven, we draw on the collated findings and conclusions to make a series of recommendations:

1. **New research** (Clist and Dercon, 2014, p1, NAO, 2015)**, available since RBA was agreed, states that agreements should see a good measure as a prerequisite for success**. This report concurs wholeheartedly, and finds RBA should not be used again in the Rwandan education sector unless better measures can be identified.

   A good measure must continue to capture something of real value (to DFID) even after it is incentivised. This point should inform both the design of the agreement, and decisions regarding verification. Specific findings that underpin this recommendation include:

   - The inability for data on the number of students sitting exams to capture learning outcomes. (Pritchett, 2013; ICAI 2015).
   - The verification exercise was unable to inform a judgement of whether gaming occurred (S3 in 2013 being the most relevant case, though we think it unlikely) **[Finding 8].**
   - The endline test of English proficiency was not consistent with the baseline, meaning aid funds were disbursed for potentially illusory success: the true nature of any improvement is unknown. This may have negatively affected policy making by highlighting a (potentially false) measured improvement and allowing communication of a *'job done'* message **[Finding 11].**

   In short, the measures used did not meet the relevant criteria. There is weak evidence of some policy effects **[Findings 10, 12]**, no convincing evidence that RBA caused greater results **[Findings 3, 4, 6, 7, 8]**, and some concerns of adverse effects **[Findings 11, 15].**

2. In the specific case of RBA to the Rwandan education sector, there are two possible avenues for future agreements. First, **RBA could use a learning indicator** (as discussed in the year two evaluation report; these are now feasible (Upper Quartile, 2015)). Second, **RBA could use English proficiency as the performance measure.** While this is not an ultimate outcome, improvement is more feasible within the kinds of timelines that DFID are likely to be working with, and, in the Rwandan context, it is a necessary condition (if not sufficient) in improving educational standards.

   Attention should be paid as to how any future agreement may influence the priorities of the GoR, and the ability of key actors to meet those priorities. In this specific case, RBA appears to have 'locked-in a pre-existing policy **[Finding 12]** whilst having an ambiguous effect on the relevant budget **[Finding 13]**.

3. **Greater attention should be paid to the political context in which PbR operates, both in future research and agreements**. In particular, the evaluation suggests that future designs consider the degree to which a measure reinforces or cuts across political incentives for the leadership of the sector, the likelihood of priorities shifting within the policy cycle, the degree of management control over the bureaucracy and how the financial disbursement is passed on to the leadership. Further research is needed to explore the effects in a wider range of political environments.

4. **Efforts should be made to ensure that in future the benefits of understanding and delivering RBA improvements outweigh the costs.** This is not simply about agreeing a tariff that is large enough, but about agreeing suitable timeframes and working with the prevailing political incentives. Crucially, it means ensuring that the financial disbursement is passed on to the sector, not lost in the budget process that requires negotiation, such as with MINECOFIN, and which would be a condition before any PbR was made. This recommendation is related to:

   - The widespread misperceptions of RBA, at even the highest levels (see section 3.3).

- That certain features (e.g. payment was to MINECOFIN not MINEDUC) were not appreciated until after the pilot came to an end (see section 3.3).

5. **Future negotiations should bear in mind that recipient governments have more information on likely improvements than donors;** this may prevent payment being made for improvements that may have happened anyway **(Finding 2** and Clist, Verschoor 2014, p.17). If improvements are not really *additional*, payments are subject to extra costs (e.g. disbursement is subject to delay and volatility) with no extra benefit **[Finding 15].** Furthermore, they may highlight data which is misleading.

5. References

Andrews, M. (2013) *The Limits of Institutional Reform in Development*. Cambridge University Press: Cambridge.

British Council (2014) *Endline Assessment of English Language Proficiency of School Teachers in Rwanda*, Revised Draft Report

British Council (2015) *Summary of the 2014 Endline Assessment of English Language Proficiency of School Teachers in Rwanda*.

Bond (2014) *Payment by Results: What it Means for NGOs.* [Online] Available at: Bond.org.uk.

Center for Global Development (2014) *12 Principles for Payment by Results (PbR) in the real world*. [Online] Available at: http://www.cgdev.org/blog/12-principles-payment-results-pbr-real-world-0.

CfBT (2015) *Capacity Strengthening for Decentralised Education Planning and Monitoring* (DEPAM) Diagnostic Report. Prepared by CfBT Education Trust.

Clist, P. and Dercon, S. (2014) *12 Principles for Payment by Results (PbR) in International Development*. [Online] Available at: http://r4d.dfid.gov.uk/pdf/outputs/Misc_Infocomm/clist-dercon-PbR.pdf.

Clist, P., and Verschoor, A. (2014) *The Conceptual Basis of Payment by Results.* University of East Anglia.

Council of Europe (undated) *Common European Framework of Reference for Languages (CEFR).* [Online] COE Language Policy Unit: Strasbourg. Available at: http://www.coe.int/t/dg4/linguistic/cadre1_en.asp [Accessed 28 September 2015]

Department for International Development (2011a) *Rwanda Education Sector Programme (RESP) 2011-12 to 2014-15: Business Case*. Department for International Development: London.

Department for International Development (2011b) DFID Ethics Principles for Research and Evaluation. Department for International Development: London. Available at: https://www.gov.uk/government/.../dfid-ethics-prcpls-rsrch-eval.pdf [Accessed 29 September 2015]

Department for International Development (2011c) *DFID's Approach to Value for Money (VfM)*. Department for International Development: London.

Department for International Development (2011d) *Implementing DFID's strengthened approach to budget support: Technical Note*. Department for International Development: London.

Department for International Development (2012) *Terms of Reference for Evaluation of Project of Results Based Aid in the Education Sector – Rwanda*. Department for International Development: London

Department for International Development (June 2014) *Sharpening incentives to perform: DFID's Strategy for Payment by Results.* Department for International Development: London.

Department for International Development and the Government of Rwanda (2012). *Results-Based Aid Pilot in the Education Sector: Annex to Memorandum of Understanding*. Kigali.

Drew, R. (July 2012). *Technical Assistance to Determine the Feasibility and Potential Uptake of an Evaluation of DFID Rwanda's Results Based Aid Pilot in the Education Sector*. Unpublished.

HEART (2014) *Independent Verification Of Educational Data For A Pilot Of Results-Based Aid (Rba) In Rwanda: Second Annual Verification Report For Student Completion Results*.

HEART (March 2015) *DFID Results-Based Aid in Rwanda; Third Annual Independent Verification Report.*

HM Government (July 2011) *Open Public Services White Paper*. London: CM 8145.

ICAI (2015) *Follow-up of ICAI Reviews from Years 1, 2 and 3* [Online]. Available at: http://icai.independent.gov.uk/wp-content/uploads/ICAI-Report-Follow-up-of-ICAI-Reviews-from-Years-1-2-and-3.pdf

National Audit Office (June 2011) "Assessing value for money" in *Successful Commissioning Toolkit*. [Online]. Available at: http://www.nao.org.uk/successful-commissioning/general-principles/value-for-money/assessing-value-for-money/

National Audit Office (June 2015) *Outcome-based payment schemes: government's use of payment by results*. London: HC 86 Session 2015-16.

Negash, N. (2011) "Teaching English, Paper 8; English language in Africa: An impediment or a contributor to development?" In *Dreams and Realities: Developing Countries and the English Language*, edited by H. Coleman. British Council: London.

OECD (2005/2008) *The Paris Declaration on Aid Effectiveness (2005); Accra Agenda for Action (2008)*. Paris: Organisation for Economic Cooperation and Development.

Pawson, R. and Tilley, N. (1997) *Realistic Evaluation*. London: Sage.

Perakis, R. and Savedoff, W. (2014) *Payment by Results: One Size Doesn't Fit All*. [Online] Center for Global Development . Available at: http://www.cgdev.org/blog/payment-results-one-size-doesnt-fit-all.

Perakis, R. and Savedoff, W. (February 2015a) *Does Results-Based Aid Change Anything?* [Online] Center for Global Development. Available at: http://www.cgdev.org/blog/does-results-based-aid-change-anything.

Perakis, R. and Savedoff, W. (February 2015b) *Does Results-Based Aid Change Anything? Pecuniary Interests, Attention, Accountability and Discretion in Four Case Studies*. Center for Global Development: Policy Paper 052, Washington D.C.

Pritchett, L. (2013) *The Rebirth of Education: Schooling Ain't Learning*, Center for Global Development: Washington DC.

Sandefur, J. and Glassman, A. (2015) The Political Economy of Bad Data: Evidence from African Survey and Administrative Statistics, *Journal of Development Studies*, 51(2), pp. 116-132.

Savedoff, W. (January 2010) *Issues in Applying the Cash on Delivery Approach*. [Online] Center for Global Development. Available at: www.cgdev.org/sites/default/files/applying-cod-aid.pdf.

Stern, S. et al. (2012) *Broadening the range of designs and methods for impact evaluations*. Department for International Development: Working Paper 38, London.

Talbot, T. and Barder, O. (August 2014) *Snakes on a Log Frame: Cobra Effects and Payment by Results.* [Online] CGD. Available at: http://www.cgdev.org/blog/snakes-log-frame-cobra-effects-and-payment-results.

Upper Quartile (2013) *Evaluation of Results Based Aid in Rwandan Education: Inception Phase Report* (unpublished).

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

Upper Quartile (2015) *Evaluation of Results Based Aid in Rwandan Education – Year Two Evaluation Report* (unpublished).

APPENDICES

Appendix 1 – Terms of Reference for Evaluation of Project of Results Based Aid in the Education Sector – Rwanda

Introduction

1. Enormous progress has been made by the Government of Rwanda in substantially expanding access to education in recent years. The government has ambitious plans to further improve access to and quality of education, articulated in its Education Sector Strategic Plan (ESSP) for 2010-2015. Development partners, including DFID, are committed to supporting the Government of Rwanda in implementing its ESSP.

2. DFID is piloting a programme of results-based aid in the education sector as part of the Education Service Delivery Grant (ESDG) of the Rwanda Education Sector Programme (RESP) which will run from 2011/12 to 2014/15. DFID Rwanda is seeking a team of consultants to conduct an impact and process evaluation of the RBA pilot.

The Objective

3. The objective of the programme is to pilot the provision of additional results-based aid based on (a) improvements in the number of students completing primary (P6), lower secondary (S3) and upper secondary (S6) education; and (b) the competency of teachers in Rwanda to use English as the means of instruction. DFID funding for the proposed RBA pilot is in addition to DFID's existing support to the education sector.

4. Key elements of the RBA pilot have been agreed between DFID and the Government of Rwanda and are summarised in Appendix 1 of this TOR.

Results Based Aid

5. Payment by results (PBR) is a new form of aid financing that makes payments contingent on the independent verification of results. PBR is part of a wider UK government agenda and several other government departments are piloting this approach. PBR is strongly referenced in the UK Government Cabinet Office's *Open Public Services White Paper*, which sets out the Government agenda for public sector reform.

6. Internationally, definitions vary - DFID makes an important distinction between those that involve payments from funders to partner governments (results-based aid – RBA - including Cash on Delivery[49]) and those that involve payments from a funder or government to service providers (results-based financing - RBF). RBA is a newer and more innovative instrument.

7. Both RBA and RBF have three key elements:

   - payments based on results;
   - recipient discretion – i.e. the recipient has space to decide *how* results are achieved[50]; and
   - verification[51] of results as the trigger for disbursement.

8. DFID has a mandate to pilot test a number of different approaches to PBR, in different sectors. These pilots are expected to focus on outcomes and to build in rigorous verification and evaluation from the beginning[52].

The Recipient

[49] This form of RBA was proposed by the Centre for Global Development, see http://www.cgdev.org/section/initiatives/_active/codaid

[50] As with all UK aid, our partnership commitments will still apply.

[51] This verification involved checking accuracy of results through quality audit processes to ensure tests are free, fair, and standards are consistent over time.

[52] DFID Primer on Results Based Approaches 2010.

9. The recipients of this service are the Government of Rwanda and DFID Rwanda.

Scope of Work

10. Given that this is a pilot, it is important that the project includes rigorous, independent evaluation. DFID and the Government of Rwanda are particularly committed to learning lessons and identifying best practice from this pilot project.

Aim

11. The aim of this piece of work is to conduct a rigorous evaluation of the pilot programme of results-based aid in the education sector in Rwanda. It is expected that the evaluation will have two main elements:
    - A process evaluation which will seek to identify the recipient's, and other key actors', response to the RBA, including establishing processes that led to any increased educational results. Among other objectives, this element of the evaluation will seek to determine any contribution made by any observed increase in the number of teachers competent to use English as the medium of instruction to any observed increase in the numbers of students completing P6, S3 and S6.
    - An impact evaluation which will seek to address whether or not the RBA pilot led to increased educational results.

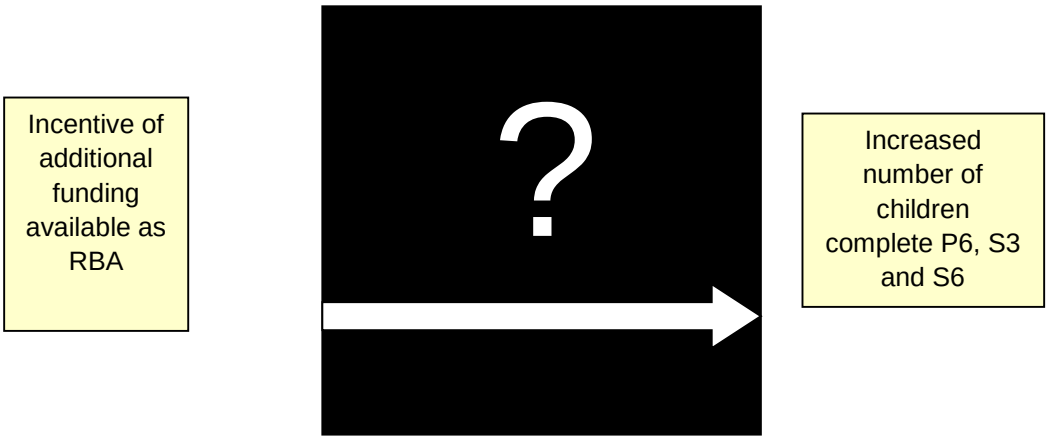
12. In addition, an annual evaluation report will be required. This annual report will serve to provide updates on progress on the evaluation. Further, the annual report will assess how the results-based aid element is working; this will allow for feedback to the design of the pilot and consequent pilot amendments.

13. These two elements of the evaluation are discussed in detail in the sections that follow. Possible questions to be asked under each of these elements are indicated in the relevant sections. However, these are indicative. It is expected that the definitive list of questions would be agreed between DFID-R, the Government of Rwanda and the selected SP through the acceptance of an inception report.

Impact Evaluation

14. The main aim of the impact evaluation element is to determine whether or not the additional incentive of results-based aid had any effect on the number of children completing different levels of education when compared with what would have been achieved without the provision of this results-based aid. This is shown diagrammatically in Figure 1. For the purpose of the evaluation, the comparison is to be between the provision of results-based aid (of up to £9m) and non-provision, with all other factors remaining constant, i.e. the counterfactual is non-provision of the RBA pilot rather than provision of a similar value of aid through another modality.

Figure A1.1: Diagrammatic representation of hypothesis that provision of additional funding as RBA results in more students completing different levels of education (P6, S3 and S6) than would have occurred without RBA payments.



15. The first step in this element will be determining if the number of children completing P6, S3 and S6 has increased. At one level, this is expected to be a simple task based on data reported by MINEDUC and verified by the process outlined above. However, there is another level of more detailed analysis of any observed increases. Relevant questions may include who benefited from these improved results? How equitable are they? Which regions benefitted more/ less? Which socio-economic groups? Boys vs girls?

16. The second step is to determine what the main drivers of the additional results were including what contribution the RBA pilot made to these additional results. For this purpose a counterfactual is required. Experimental and quasi-experimental designs (including phasing designs or pilots at sub-national level) are unlikely to be feasible for the programme due to the barrier of not being able to establish a control group (the pilot is being rolled out nationally).

17. DFID expects the SP to adjust the impact evaluation in response to feedback from the recipient; DFID is open to suggestions about the most appropriate approach and methodology to establish causality in this context. One potential approach that would allow causality to be established is construction of a counterfactual through a prospective mathematical model to predict expected results without the additional results-based aid element. The SP would then be expected to compare actual observed results obtained from all planned elements with inclusion of results-based aid with the mathematical model to see if the observed results fall within or outside the model's parameters. The SP would then analyse these observations to draw conclusions on causal inference between the RBA pilot and the observed results, that is the evidence of whether or not the additional incentive of results-based aid led to a greater level of results than would have been achieved in the absence of this incentive.

18. DFID is open to other appropriate approaches and methodologies to establish a counterfactual, but those identified must be sufficiently robust to allow for causality to be established. In 2012, DFID published a study entitled 'Broadening the Range of Designs and Methods for Impact Evaluation, DFID Working Paper 38' (Stern, E. et al). This study sets out approaches for assessing impact that do not rely on experimental approaches. The study can be found on the DFID website at the following link - http://www.dfid.gov.uk/r4d/Output/189575/Default.aspx

19. The SP should also assess the level of results-based aid on offer with respect to the strength of incentive that it creates. For example, an assessment should be made of whether the results-based aid element creates an adequate incentive for improved results. Other questions include: 'Is DFID paying for results that would have been achieved anyway?' 'Are the results stretching enough/too stretching?' The annual evaluation report should recommend whether the thresholds should be reassessed (up or down) in light of the evidence generated by the evaluation in the previous year.

Process Evaluation

20. The evaluation is also expected to examine the response to the RBA, including the mechanisms through which results-based aid led to any increase in results. This essentially involves 'unpacking the black box' shown in **Figure A1.1.**

21. The SP should use the current RESP theory of change model as a framework for this 'unpacking' **(Figure A1.2)** as this reflects the Government of Rwanda's thinking on how educational results will be delivered in the country. This has been slightly modified from the version presented in the DFID Business Case to label different levels (0-3) of the results chain and to highlight the importance of teachers' competency to use English as the medium of instruction. Tenders which are not based on this theory of change and/or propose alternate theory of change models will not be considered. The evaluation is expected to assess the extent to which observed changes have occurred as a result of the processes outlined in the RESP theory of change and/ or have occurred as a result of other processes not captured in that model. The evaluation is expected to critically assess the processes by which educational results, such as an increase in the number of students taking examinations, occurs in Rwanda and the extent to which the RESP Theory of Change reflects these processes.

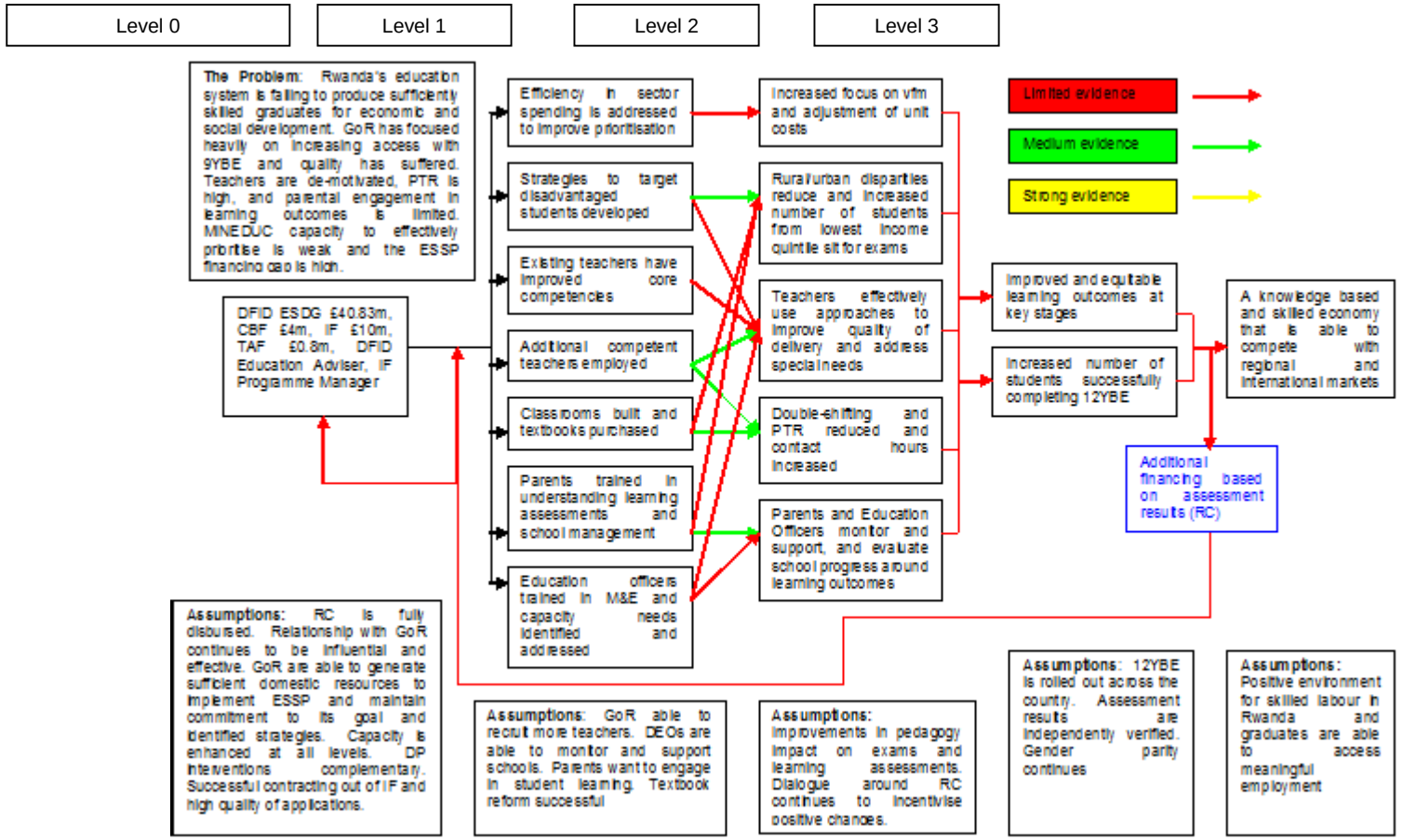
22. Additionally, the SP should gather evidence of the extent to which RESP, in general, and the results-based aid pilot, in particular, have contributed to different elements of the results chain and the extent to which these elements have led to others. Indicative examples of the types of questions that might be asked by the evaluators at different levels of the results chain are presented in **Appendix A1.2.** However, bidders may suggest revisions, adjustments or additions to these proposed questions. It is expected that this list of questions would be agreed between DFID-R and the SP through the acceptance of an inception report.

23. It is expected that the SP will also pay particular attention to identifying how any observed changes in the number of teachers competent to use English as the medium of instruction have occurred. For example, this would involve unpacking the extent to which any improvement is as a result of improving the English language skills of existing teachers or as a result key factors including recruiting new teachers already competent in English and the school mentoring programme. Although this approach would involve analysing the contribution made by the school mentoring programme, it is not expected that SP's would conduct a detailed evaluation of that programme.

24. The evaluation team will also be expected to identify any unexpected consequences of RBA including, in particular, any perverse incentives created specifically by RBA.

25. As with the impact evaluation, DFID-R is flexible on the choice of approach and methodologies to underpin the process evaluation. DFID-R considers, however, that there is merit in pursuing a realist evaluation approach. A realist evaluation seeks to collectively understand if an intervention works, (and if the intervention works) for whom the intervention works and under what circumstances the intervention works. In addressing these points, it is critical to explore the context (including economic, social, political, cultural, and historic background, and organisational set-up including resources supporting the intervention). Further, a realist evaluation seeks to explain on a dynamic basis the interaction of context and the intervention, and to test the likely effectiveness of alternative contexts and interventions. Realist evaluation is particularly useful for informing policy, due to the ability to apply findings to other settings.

Figure A1.2: Proposed RESP Theory of Change (modified from DFID Business Case[53])



[53] The level labels are not part of the original diagram and have been added for ease of reference in these terms of reference.

26. In addition, the evaluation is expected to explore:
    - Whether the provision of RBA and the focus on increasing the number of children completing particular levels of schooling had any positive or negative effects on equity issues. Equity issues that the evaluation is expected to explore include whether results-based aid disproportionately benefited:

        i. One sex more than another?
        ii. Any geographical area more than another?
        iii. Those children from higher wealth quintiles?

    - Any effect on aid relationships. This element of the evaluation should cover any effects of the RBA pilot on relationships between GoR and DFID, and between GoR, DFID and other development partners. The SP should focus particularly on interactions between the RBA pilot and DFID's policy dialogue with GoR. For example, 'in what way did DFID's policy dialogue contribute to or hinder results achieved?' 'In what way did the RBA pilot increase or reduce DFID's policy dialogue with GoR?'

    - Any effect on accountability to citizens. An important theoretical aspect of RBA is its proposed ability to promote citizen empowerment and accountability, for example, through the transparent publication and dissemination of results. The evaluation is expected to explore the extent to which publication has happened and how any observed change was achieved.

Evaluation Scope

27. It is expected that the evaluation would have an initial inception phase in which the SP would:
    - Outline their understanding of the evaluation, highlighting how they would address key issues and overcome limitations, in order to ensure that the aim of the evaluation is achieved within the timeframe
    - Begin development of the mathematical model or other, alternate approach proposed to be used to generate the counterfactual
    - Finalise the method and approach to be used for the process evaluation
    - Finalise the indicators to be tracked and the questions to be asked in the process evaluation
    - Finalise the timeline and workplan

28. One question that arises in relation to this evaluation is the extent to which the evaluators are being expected to evaluate the whole of RESP or just the RBA component. In principle, the evaluation is focused on the RBA component only. However, as the RBA component is embedded within RESP, in general, and ESDG, in particular, the evaluation will need to make some overall assessment of these instruments.

29. It is expected that the evaluation will adhere to OECD DAC evaluation criteria and standards.

Method

30. For the impact element of the evaluation, the SP will be expected to develop a model for the counterfactual and to compare the verified results reported with expected results generated by this model. This could include mathematical modelling or other robust approaches of establishing impact. Tenders which do not propose a robust approach to impact evaluation and do not include team members with skills in this area will not be considered. It is expected that proposed methods and approaches will be elaborated / finalised during the inception phase.

31. For the process element of the evaluation, it is expected that the method and approach to be used will be finalised during the inception phase. Bidders are however expected to propose methods and approaches in their tenders. Final choice of method to be elaborated at inception and will depend on:

    - The indicators to be tracked and the questions to be asked.
    - The extent to which relevant primary data is available from MINEDUC.

32. In principle, DFID would not expect the SP to have to do large surveys or significant amounts of primary data collection for the process evaluation. However, this would not exclude conducting some primary data collection in focused areas, such as surveys or focus group discussions to assess the degree of citizen empowerment and the perceptions of pupils and parents. This should only be done where it would not be possible/ appropriate for this to be done through national systems. In general, DFID's expectation is that the majority of primary data collection would be done through national data systems, such as EMIS and the system for examinations data. It is expected that data related to the competence of teachers to use English as a medium for instruction would be collected by the Government of Rwanda through surveys of representative samples of teachers in both 2012 and 2014.

Use of Evaluation Findings

33. DFID expects that lessons learned from the evaluation will be used by MINEDUC, DFID and other development partners in a range of different ways, particularly to further improve the education sector in Rwanda and to shape DFID's policy on results-based aid more broadly. Bidders are expected to explain how they would support the process of lesson learning if appointed, including through the provision of policy relevant advice.

34. Also, DFID expects that the SP will include clear and concise recommendations on key lessons concerning the processes and approaches used to achieve intended and unintended results through the RBA incentive.

35. DFID expects the final evaluation report to be formally published, for example, in a peer-reviewed journal. Bidders are expected to outline how they would ensure this in their proposal including suggestions as to where/how this should be done.

Reporting Requirements

36. It is expected that the evaluation will produce the following deliverables against which payments will be made. Timings are based on the assumption that a SP will be in place by end December 2012

A written inception report within six months of contract signing and submitted to DFID-Rwanda. This report would be expected to present the approach to be used for the impact evaluation including the proposed counterfactual; finalise the method and approach to be used for the process evaluation; finalise the indicators to be tracked and the questions to be asked in the evaluation; and finalise the timeline and workplan.

- Annual reports on the progress of the evaluation, updating baseline data and making recommendations for any adjustments to the project design and implementation. It is proposed that the first written reports should be produced within 12 months of contract signing with a follow up annual report each year.

- A draft and final evaluation report covering the whole evaluation period which presents evaluation findings, challenges and lessons learnt with clear recommendations to DFID, MINEDUC and other stakeholders relating to the design and implementation of results-based approaches in the education sector. This should be no more than 30 pages, excluding annexes and supplementary material. The draft written report is due to be submitted to DFID Rwanda by the 20th May 2015 with the final report due on or before the 30th June 2015.

37. The SP will report through the DFID Education Advisor to the RBA Evaluation Steering Committee, comprised of the Government of Rwanda and DFID Rwanda.

38. The Steering Committee consisting of MINEDUC officials, DFID and an external expert will be constituted ensure the independence of the evaluation, provide technical guidance, address any contentious issues and discuss progress. The Steering Committee will also ensure an effective Communications Strategy is in place to guide communications in relation to the Evaluation process and outcomes. Finally the Steering Committee will undertake quality assurance to ensure technical rigour of deliverables.

Suggested Expertise

39. This consultancy requires a small core team of international experts supported by a small team of national experts. Precise team composition can be proposed by bidders. Staff numbers and cost should be proportionate to the overall size of the RBA project.

40. DFID expects that the team leader would be an evaluation specialist with experience of conducting evaluations of this nature with elements of impact and process evaluation. Experience of the education sector is not considered essential for the team leader. Rather, DFID expects a team leader with high levels of evaluation expertise.

41. DFID also expects that one team member would be an expert in mathematical modelling/ alternative approaches to establish causality with experience of creating counterfactuals for the basis of evaluation. Experience of the education sector is not considered essential for this expert. Rather, DFID expects that this team member has high levels of impact evaluation expertise.

42. DFID envisages some aspects of the evaluation being undertaken by a Rwandan partner. This is in line with priorities of sustainability and enhancing local capacity. Preference will be given to bids which will demonstrate that they will build the capacity of Rwandan nationals to undertake evaluation exercises.

46. DFID intends to manage the provider's performance through a suite of key performance indicators. The draft suite of indicators is contained is contained Appendix 3 of this TOR and tenderers are welcome to propose additional or alternative indicators. These indicators will be agreed after the inception phase and ultimately be incorporated in to the contract.

Background

The RBA programme forms part of DFID's Rwanda Education Sector Programme (RESP). RESP runs from 2011/12 to 2014/15 and is worth more than £55m. It includes an Innovation Fund of £10m, a Capacity Building Fund of £4m and a Technical Assistance Fund of £0.8m. The largest component of RESP is an Education Service Delivery Grant (ESDG) of £40.83m.

The ESDG is made up of two parts. The largest part will consist of £31.83m of sectoral budget support. In addition, the UK will provide up to an additional £3m per year in the financial years 2013/14, 2014/15 and 2015/16 based on achievement of agreed results above currently-expected levels. This element is termed "results-based aid" and is considered to be an innovative way of providing development aid. Any funds payable as results-based aid will be paid as additional sectoral budget support.

The RBA component will be paid in annual tranches of up to £3m per year in UK financial year 2013/14, 2014/14 and 2015/16, based on the number of students completing various stages of schooling (P6, S3 and S6) above the 2011 baseline. Taking a national examination will be used as an indicator of having completed a particular level of education. Payments will be made based on the number of students sitting the examination the previous year above the 2011 baseline multiplied by an agreed unit price, subject to the annual ceiling of £3m. For each examination level (P6, S3 and S6), there will be two different unit prices - a higher price for each additional student above previous year's levels and a lower price for each additional student above the 2011 baseline but below the previous year's performance.

An additional payment will be made in 2015, based on the number of teachers in 2014 with improved English language competency over a 2012 baseline. An independent verification will be undertaken to ensure the accuracy and reliability of data being used as the basis for results-based aid payments.

The main purpose of the evaluation is to determine the extent to which the RBA led to increased levels of results in comparison to what would have happened had the RBA not been provided. In addition, the evaluation is expected to learn key lessons about the processes and approaches used to achieve the observed results.

The same financial incentive will apply to all students completing levels of schooling regardless of gender or geographical location. However, the evaluation will be expected to explore the effect of results-based aid on equity issues.

It is proposed that any remaining funds after RBA payments have been made will be retained in Rwanda but used in a sector other than education. The evaluation should explore any effects that this retention of funds in country has on the RBA incentive for the Government of Rwanda (See **Appendix A1.2,** Level 0, Question 1).

Results-Based Aid Pilot in the Education Sector: Proposed Annex to Memorandum of Understanding

Parties and purpose

This annex outlines key, agreed elements of a pilot of results-based aid (RBA) in the education sector in Rwanda. In this pilot, DFID will make additional Sector Budget Support payments in 2013-2015 to the Government of Rwanda for results achieved in academic years 2012, 2013 and 2014. This is referred to as a Results Compact in the DFID Rwanda Education Sector Programme Business Case (2011-2015). This annex guarantees that DFID will make a fixed payment for each additional unit of progress towards educational outcomes, as stipulated below.

The arrangements under which the Grant will be disbursed are set out in the attached MoU and DFID's Partnership Commitments. The Government of Rwanda will decide on the use of any funds received.

Term of agreement and possibility of renewal

This agreement is from date of signature until May 2015. There is a possibility of renewal and/or expansion of the programme depending on the results of the pilot. Any renewal or expansion would require the agreement of both DFID and the Government of Rwanda.

In extreme circumstances, if DFID is concerned that the provisions of this agreement, or partnership commitments made under the arrangement may not have been fulfilled by the Government of Rwanda or if any changes occur which significantly impair the development value of this project/ programme, DFID will discuss with the Government of Rwanda and where appropriate undertake assessment. If warranted, such an assessment could lead to cessation of this agreement.[54]

Results and indicators

The main results to be rewarded in this pilot will be the number of children completing key stages of 12 year basic education, namely year 6 primary (P6), year 3 secondary (S3) and year 6 secondary (S6). These results will be measured by the number of children taking the P6, S3 and S6 examinations annually. These results should include all students taking each of these examinations for the first time regardless of the sector in which they are learning, i.e. public or private. However, students who are retaking an examination should be excluded from the figures of those taking the examination in a particular year. Payments will be made for any results achieved above 2011 levels. Payments will only be made after independent verification of the results. Payments will be made as set out in section D and will be an additional DFID contribution to the Government of Rwanda's efforts to meet ambitious education targets as articulated in the Education Sector Strategic Plan (ESSP).

In addition, it is agreed that an additional one-off payment will be made in 2015 based on results achieved to improve the English language competency of teachers in the education sector. It is agreed that this payment would be based on assessing the English language skills of a representative sample of Rwandan teachers at baseline (2012) and in 2014, as outlined in the payment schedule below. The level of competence in English required for a primary and secondary school teacher will be agreed by Government of Rwanda in consultation with DFID and will be set out as a further annex to this agreement. Data from this assessment would be used to calculate the number of teachers in Rwanda

[54] More details of situations of this nature are provided in the main body of this memorandum of understanding.

achieving an agreed level of competency in English. DFID will then pay the Government of Rwanda an agreed amount for every additional teacher with that level of competency in English.

Payment schedules

It is agreed that DFID will pay the Government of Rwanda additional sums up to a ceiling of £9in the three year period 2013-2015. Payments will be made annually in 2013, 2014 and 2015 based on verified results of the previous year's exams, i.e. 2012, 2013 and 2014. RBA payments will be made no later than April/May each year.

SBS and RBA Disbursement schedule

GoR FY	2011/12	2012/13	2013/14	2014/15
SBS	£8.57m	£6.37m	£7.72m	£9.17m
RBA		Up to £3m	Up to £3m	Up to £3m

Payments will be based on independently verified data, as set out in section F below. It is agreed that payments will be calculated as follows.

- For each additional child sitting the P6 exam above the previous year's results, DFID will pay the Government of Rwanda £50. In addition to this payment, in years 2014 and 2015, DFID will also pay the Government of Rwanda £10 for each additional child sitting the P6 examination above 2011 levels.
- For each additional child sitting the S3 exam above the previous year's results, DFID will pay the Government of Rwanda £100. In addition, in years 2014 and 2015, DFID will also pay the Government of Rwanda £10 for each additional child sitting the S3 examination above 2011 levels.
- For each additional child sitting the S6 exam above the previous year's results, DFID will pay the Government of Rwanda £50. In addition, in years 2014 and 2015, DFID will also pay the Government of Rwanda £10 for each additional child sitting the S6 examination above 2011 levels.

So for example, if 77,473 students took the S3 exam in 2011 and 85,000 take it in 2012 DFID would pay the Government of Rwanda (85,000-77,473)*£100 = £752,700 in 2013. If 93,000 students then took the S3 exam in 2013, DFID would make two payments to the Government of Rwanda in 2014, namely (93,000-85,000)*£100 = £800,000 plus (85,000-77473)*£10 = £75,270. This would be a total of £875,270. More details of the calculations involved are available in an Excel calculator developed by DFID.

In addition, it is agreed that in 2015 DFID will also pay Government of Rwanda £50 per additional teacher competent to use English as the medium of instruction. This will be based on a baseline assessment conducted by the Government of Rwanda in 2012 and a follow-up assessment conducted by Government of Rwanda in 2014. Any payment due would be made in 2015 based on independently verified results and subject to available funds within the £9m three year ceiling as specified in paragraph 6 of this annex.

Payment levels can only be changed with the express written agreement of both DFID and Government of Rwanda. DFID and the GoR will meet to review programme related impact, targets and costs immediately after results have been verified and the annual evaluation report received.

Use of funds

The funds that will be provided by DFID through the RBA pilot can be used as desired by the Government of Rwanda. DFID will not provide any restrictions for the use of these funds in accordance

with the principles of results based aid55. It is expected that these will be used to further improve the results being tracked, namely the number of pupils completing key education levels – P6, S3 and S6. Any decisions as to how these funds will be used rests solely with the Government of Rwanda. In line with the overall provisions of this memorandum of understanding, DFID retains the right to access audited financial statements, prepared by the Government of Rwanda, to verify that the income received has been declared and used to support the country expenditure.

Data verification, citizen empowerment, transparency and accountability

It is essential that data used to trigger payments is accurate and reliable. It is therefore necessary for the data reported for payment purposes to be verified independently. Payments will only be made on the basis of independently-verified results. DFID will hire an external contractor to conduct this work. DFID will select the contractor in consultation with the Government of Rwanda. Government of Rwanda agrees to cooperate fully with the work of this contractor which will involve checking the systems for collecting and reporting P6 to S6 exam participation rates and checking a data sample. The external contractor will also verify the Government of Rwanda baseline and end assessment of teacher competence in English language. Data verification needs to be both robust and timely. Government of Rwanda will provide DFID and designated verification and evaluation teams with full access to any necessary data required to validate results achieved. Both DFID and Government of Rwanda recognise and agree that if issues are identified in the data verification process, this may result in funds being delayed and/or withheld.

In line with DFID's Transparency commitments, the Government of Rwanda gives consent for this arrangement, and any subsequent amendments, to be published on DFID's website. The Government of Rwanda also agrees to make this agreement and the annual results of the exercise publically known.

It is essential that the design of the Results-Based Aid programme be communicated to schools and parent teacher associations (PTAs). Government of Rwanda agrees to ensure that this happens. This will ensure that teachers and parents are aware that the education sector will receive additional funds based on increasing the number of students taking P6, S3 and S6 examinations and the English language competency of teachers.

Evaluation and lesson learning

As this is a programme to pilot an innovative way of providing aid, both DFID and Government of Rwanda agree that it is essential that lessons are learned from this process. This will be done through a rigorous evaluation. The evaluation will seek to determine the extent to which the results-based aid has had an effect additional to what would have happened without it.

In addition, both DFID and the Government of Rwanda are committed to learning lessons about the processes followed to achieve the expected results. This will be done through a rigorous process evaluation based on a causal chain leading from the inputs and processes to expected outputs and outcomes. Indicators and/or evaluation questions will be identified for each of these elements/levels. (These process indicators will not be used as a basis for payments but will be used solely for learning purposes). The evaluation will also explore unexpected consequences of the results-based aid programme. Government of Rwanda agrees to cooperate fully with any evaluations of the RBA pilot project including through allowing access to data that the evaluation team requires and allowing the evaluation team access to MINEDUC/REB staff, schools, teachers and students.

Management arrangements

It is expected that there will be issues which arise in the course of operations of the RBA pilot which will require discussion, dialogue and resolution. These will be handled through the routine DFID/PS forums which take place on a monthly basis.

Amendment, dispute resolution and termination

[55] See for example DFID (2010) *Primer on results based aid and results based financing* and Birdsall, Savedoff and Mahgoub (2011) *Cash on Delivery: A New Approach to Foreign Aid*

Amendments to this arrangement need to be agreed by both governments in writing.

In case of dispute arising, attempts will be made to resolve these through the regular meeting between DFID and the Permanent Secretary of the Ministry of Education.

This arrangement may be terminated by three months written notice from either government. Any decision of either government regarding termination of this Arrangement will first be subject to discussion.

Duty of Care (DoC)

The Supplier is responsible for the safety and well-being of their Personnel (as defined in section 2 of the Framework Agreement) and Third Parties affected by their activities under this Call-down Contract, including appropriate security arrangements. They will also be responsible for the provision of suitable security arrangements for their domestic and business property.

DFID will share available information with the Supplier on security status and developments in-country where appropriate. DFID will provide the following:

- All Supplier Personnel will be offered a security briefing by the British High Commission & DFID on arrival. All such Personnel must register with their respective Embassies to ensure that they are included in emergency procedures.
- A copy of the DFID security briefing notes (and a further copy each time these are updated), which the Supplier may use to brief their Personnel on arrival.

The Supplier is responsible for ensuring appropriate safety and security briefings for all of their Personnel working under this Call-down Contract and ensuring that their Personnel register and receive briefing as outlined above. Travel advice is also available on the FCO website and the Supplier must ensure they (and their Personnel) are up to date with the latest position.

FCO advises against all but essential travel to within 1km of the border with DRC and Burundi, except to the towns of Gisenyi (Rubavu) and Cyangugu (Rusizi), and major border crossings to Burundi. This is because of instability in those countries' border areas. From 22 October, the border crossings between Rwanda and the DRC will be closed from 18:00 hours until 06:00 at Bukavu and Goma. It is likely these restrictions will remain in place for the foreseeable future. The FCO advises against travel to parts of DRC and Burundi. The areas close to the borders with Uganda and Tanzania are less dangerous but we advise travellers to keep to main roads and use recognised border crossings. The Supplier will not be required to work in these areas against FCO advice.

Tenderers must develop their Tender on the basis of being fully responsible for Duty of Care in line with the details provided above and the initial risk assessment matrix developed by DFID (see Appendix 1 of this ToR). They must confirm in their Tender that:

- They fully accept responsibility for Security and Duty of Care.
- They understand the potential risks and have the knowledge and experience to develop an effective risk plan.
- They have the capability to manage their Duty of Care responsibilities throughout the life of the contract.

Further information on Duty of Care is provided in the Supplier Instructions (Volume 1 of the Mini-Competition Invitation to Tender Pack).

TOR Appendix A1.1 - Summary risk assessment matrix

Project/intervention title: PO – **Evaluation of Project of Results Based Aid in the Education Sector – Rwanda**

Location: Rwanda

Date of assessment: 06/09/2012

Assessing official: Sifa Uwera

Theme	DFID risk score
OVERALL RATING[56]	2
FCO travel advice	3
Host nation travel advice	Not available
Transportation	2
Security	2
Civil unrest	1
Violence/crime	2
Terrorism	3
War	1
Hurricane	1
Earthquake	3
Flood	2
Medical services	3
Nature of project/intervention	2

1=very low risk	2= low risk	3=medium risk	4=high risk	5=very high risk
Low		Medium	High	

[56] The overall risk rating is calculated using the MODE function which determines the most frequently occurring value

TOR Appendix A1.2 - Possible evaluation questions at different levels of the RESP results chain

	Results chain element	Possible evaluation questions
	LEVEL ZERO	
1	RESP financing	Are there any issues relating to the way DFID provides funds which have affected delivery of RBA pilot? In particular, (i) have there been any positive or negative effects of having an annual financial ceiling; (ii) have there been any positive or negative effects of the way in which funding left over from RBA has been used; (iii) have there been any positive or negative effects of having fixed, annual tranches available rather than a variable amount based solely on results achieved?
	LEVEL ONE	
1	Efficiency in sector spending is addressed to improve prioritisation	What evidence is there that (i)RESP (ii)RBA has resulted in (a) more efficient sector spending (b) improved prioritisation? What effect (if any) has there been on value for money?
2	Strategies to target disadvantaged students developed	What strategies does MINEDUC have to target disadvantaged students? How has (i) RESP (ii) RBA contributed to these?
3	Existing teachers have improved core competencies	To what extent have the core competencies of teachers improved, e.g. in using English as the medium of instruction? How has (i) RESP (ii) RBA contributed to these?
4	Additional competent teachers employed	How have patterns of teacher employment changed over the lifetime of RESP? Is there evidence that recently-employed teachers have higher levels of competency, e.g. in using English as the medium of instruction? How has (i) RESP (ii) RBA contributed to these?
5	Classrooms built and textbooks purchased	How many classrooms have been built? How many textbooks purchased? How has (i) RESP (ii) RBA contributed to these?
6	Parents trained in understanding learning assessments and school management	How many parents have been trained to understand learning assessments? How many parents have been trained in school management? How has (i) RESP (ii) RBA contributed to these?
7	Education officers trained in M&E and capacity needs identified and addressed	How many education officers trained in M&E? What capacity needs were identified? How were these addressed? How has (i) RESP (ii) RBA contributed to these?
	LEVEL TWO	
1	Increased focus on value for money and adjustment of unit costs	What evidence is there of an increased focus on value for money? What has happened to unit costs? To what extent do these changes reflect (a) increased efficiency in sector spending (b) improved prioritisation
2	Rural/urban disparities reduce and increased number of students from lowest income quintiles sit for exams	Are there any differences/trends in the results observed in terms of children completing different levels of education for (a) boys vs girls; (b) rural vs urban settings; (c) children from highest and lowest wealth quintiles? To what extent are any changes due to: (i) specific strategies to target disadvantaged students; (ii) patterns of classroom construction/textbook distribution; (iii) patterns of training parents?
3	Teachers effectively use approaches to improve quality of delivery and address and address special needs	What evidence is there of teachers effectively using approaches to (i) improve quality of delivery (ii) address special needs. To what extent are teachers in Rwanda able to use English as a medium of instruction? To what extent are any changes due to (i) specific strategies to target disadvantaged students;(ii) teacher training in learner centred methodologies; (iii) more teachers employed and better remuneration; (iv) better trained education officers

	Results chain element	Possible evaluation questions
4	Double shifting and PTR reduced and contact hours increased	To what extent has (a) double shifting reduced; (b) pupil teacher ratio reduced; (c) contact hours increased? To what extent are any changes due to (i) more teachers being employed and being better remunerated; (ii) more classrooms being built and more textbooks being distributed?
5	Parents and Education Officers monitor and support, and evaluate school progress around learning outcomes	To what extent do (a) parents; (b) Education Officers (1) monitor; (2) support and (3) evaluate school progress around learning outcomes? To what extent are any changes due to (i) parent training; (ii) Education Officer training?
	LEVEL THREE	
1	Improved and equitable learning outcomes at key stages	What evidence is there of improved learning outcomes? How equitable are these? To what extent are any changes due to (i) increased focus on value for money and adjustment of unit costs; (ii) addressing inequalities between (a) rural/urban areas;(b) lowest and highest wealth quintiles; (iii) teachers more effectively using approaches to improve quality of delivery and address special needs; (iv) reduction of double shifting, reduction of PTR and increasing contact hours; (v) increased monitoring, support and evaluation from (a) parents and (b) Education Officers?
2	Increased number of students successfully completing 12YBE	What has happened to numbers of students completing different stages of 12YBE? P6? S3? S6? To what extent are any changes due to (i) increased focus on value for money and adjustment of unit costs; (ii) addressing inequalities between (a) rural/urban areas;(b) lowest and highest wealth quintiles; (iii) teachers more effectively using approaches to improve quality of delivery and address special needs; (iv) reduction of double shifting, reduction of PTR and increasing contact hours; (v) increased monitoring, support and evaluation from (a) parents and (b) Education Officers?

TOR Appendix A1.3 - Generic Key Performance Indicators

DFID Key Performance Criteria	Sub Criteria
	How do you rate performance against:
Quality & Delivery	Quality and timeliness of milestones/deliverables
	Quality of monitoring, evaluation and reporting
	Project impact / key results on track according to programme logframe
Management, Strategy & Financial	Monthly and quarterly reports submitted to agreed deadlines
	Grants are disbursed within agreed deadlines
	Ability to deliver in line with agreed budget
Personnel	Performance of team leader
	Performance of other team personnel
	Performance of country based teams
	Timeliness in replacing personnel with appropriate level of approval from DFID
	Managing underperformance
Customer Relationship	Risk Management
	Responsiveness to stakeholders
	Impact of outreach and external communications
	Regular communication with DFID and delivery of weekly meeting action points
	Development of new delivery partners
	Attention to DFID policies
Innovation and Continuous Improvement	Provider has sought to improve on the last reporting periods performance
	Provider has delivered new ideas which have benefited the programme
Corporate Social Responsibility	Activities have resulted in a positive effect on sustainable/environmental considerations
	Supply Chain: the amount of local contractors used within the supply chain to deliver the programme
	Employment: Apprenticeships, local opportunities

Overall Performance in terms of Value for Money

Appendix 2 – Terms of reference for technical assistance to explore options for assessing value for money of the Results Based Aid pilot in the Education Sector

A. Summary

Upper Quartile are being asked by DFID Rwanda to conduct additional work on options for assessing value for money of the Results Based Aid (RBA) pilot in the Education Sector as part of their ongoing work of evaluating the RBA pilot (see UQ and IPAR, 2014).

B. Context

The UK's Coalition Government has set out a clear agenda for its development aid focused on results, value for money and accountability. As part of this, DFID is piloting a number of RBA programmes. The idea behind RBA is that the donor agrees to pay for the achievement of certain results but recipients are empowered to decide how those results should be achieved and delivered.

In Rwanda, DFID is piloting a programme of results-based aid in the education sector as part of the Education Service Delivery Grant (ESDG) of the Rwanda Education Sector Programme (RESP) which will run from 2011/12 to 2014/15. The objective of the programme is to pilot the provision of additional results-based aid based on (a) improvements in the number of students completing primary (P6), lower secondary (S3) and upper secondary (S6) education; and (b) the competency of teachers in Rwanda to use English as the means of instruction. DFID funding for the proposed RBA pilot is in addition to DFID's existing support to the education sector.

DFID is currently supporting an evaluation of this pilot which has two main elements:

- A process evaluation which is seeking to identify the recipient's, and other key actors', response to the RBA, including establishing processes that led to any increased educational results. Among other objectives, this element of the evaluation will seek to determine any contribution made by any observed increase in the number of teachers competent to use English as the medium of instruction to any observed increase in the numbers of students completing P6, S3 and S6.
- An impact evaluation which is seeking to address whether or not the RBA pilot has led to increased educational results.

This evaluation is being implemented by Upper Quartile. Although there was some expectation that the evaluation would consider the value for money of the RBA pilot, this was not originally anticipated to be a major area of focus for the evaluation. However, it has emerged during implementation that this is a crucial area for the pilot which needs further exploration and this is the reason why DFID is requesting additional work which will be incorporated as a contract amendment.

C. Purpose and objectives

The purpose of this assignment is to explore options for assessing the value for money of the Results Based Aid pilot in the education sector.

The assignment has two objectives. The first objective is considered the main or primary objective and it is expected that most of the assignment's focus would be on this objective. The second objective is considered a subsidiary or secondary objective. The assignment is expected to conduct some work on this objective. However, it is recognised that such work may be preliminary and that it is likely to result in identification of further work that might need to be done.

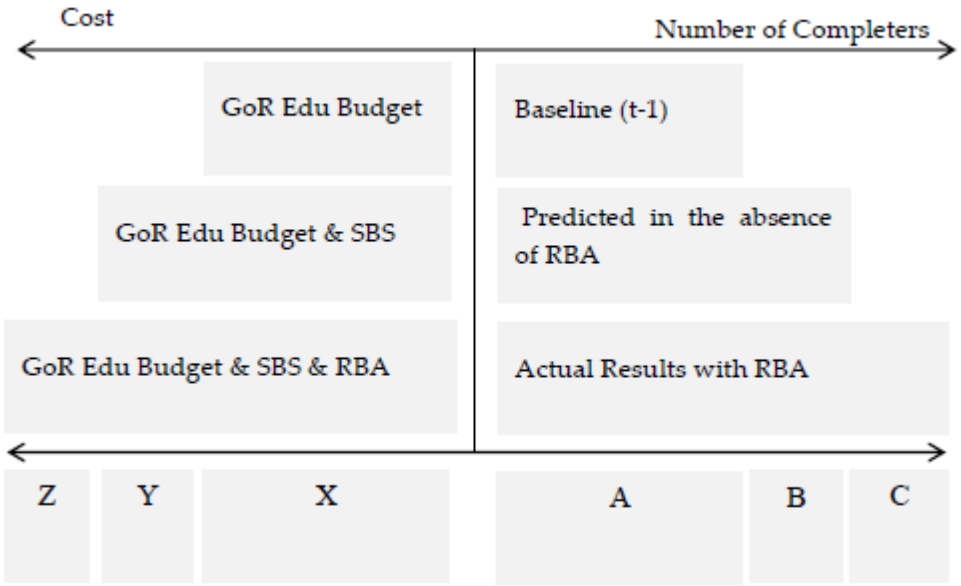
The assignment's first/main objective is to consider and present data for four options for assessing cost-effectiveness as outlined in the evaluation team's Approach Paper on Value for Money in RBA, Rwanda

(Clist, 2014). More details of these options are contained in that paper but these are presented in brief in **Table A2.1** and **Figure A2.1.**

Table A2.1: Options for assessing cost effectiveness of RBA pilot in the Education Sector in Rwanda (for more detail, see Table 1 in Clist, 2014)

Option	Benefits	Costs	Comparison[57]
1	(Z+Y)/(X+Y+Z)(A+B+C)	(Z+Y)	Zero
2	B+C	Z	SBS where benefits are A and cos are X+Y
3	B+C	Z	Zero
4	C	Z	Zero

Figure A2.1: Three levels of costs and benefits (from Figure 2 in Clist, 2014)



In **Figure A2.1:**

- X is the Government of Rwanda education budget including GoR funds and all other development partner funding but excluding sectoral budget support and RBA from DFID
- Y is DFID's sectoral budget support in education
- Z is the cost of results-based aid including cost of verification and management (but excluding the cost of evaluation)
- A is the baseline level of completers, that is historic data from 2011
- B is the predicted additional performance above baseline based on the econometric model developed by the evaluators
- C is the observed results with RBA above the level predicted by the evaluators' econometric model

The contractor is expected to clarify in an updated VFM approach paper whether:

- Costs will be calculated on an annual basis, i.e. across the education sector, as stated in the approach paper, or whether they will be calculated on a multi-year basis, i.e. for the cohort completing the level of education specified
- Benefits will be calculated on completion of each of the three levels rewarded by RBA (P6, S3 and S6) or whether the assessment will focus on only one of these levels.

[57] Two comparisons are envisaged – a comparison with no aid and a comparison with existing aid.

In calculating the benefits from education, the contractor will be expected to factor in health benefits of completing particular levels of education. The contractor should also consider is other benefits of education can be factored into the calculations. These include delayed marriage for girls, better health of children of girls, citizen engagement. More details of these are available in DFID's education position paper.

The assignment's second/subsidiary objective is to consider how the economy, efficiency and effectiveness of results based aid might be assessed including the usefulness or otherwise of:

- Measuring the level of transaction costs, particularly of data verification, and comparing these with the transaction costs of other aid modalities.
- Calculating unit costs, i.e. total cost of each additional completer (including management costs. Is it possible assess the extent to which the price level set was optimal?

D. Scope of the work

The contractor will be expected to liaise with similar work being carried out to assess value for money of DFID's RBA pilot in the Education Sector in Ethiopia. However, the contractor will not be expected to calculate value for money for that programme. If figures for value for money are available for that programme, the contractor would be expected to compare and contrast those figures with those that they generate for Rwanda. The contractor is not expected to formally compare value for money of the RBA pilot in the education sector in Rwanda with DFID funding of other sectors in Rwanda, or funding of other agencies to education in Rwanda, or DFID funding to education in other countries (with the exception of the DFID RBA pilot in education in Ethiopia referred to above). However, suggestions as to how such comparisons could be done could, if feasible, be included in the contractor's final report.

The contractor is not expected to assess the economy, efficiency or effectiveness of how the Government of Rwanda uses funds provided by DFID through RBA.

The contractor is not expected, as part of this additional work, to assess whether or not results based aid achieved additional results or paid for results that would have occurred anyway as this is already a key element of the RBA evaluation. However, the contractor would be expected to use this information in its value for money calculations as outlined above.

E. Method

The bidder is expected to outline briefly in their proposals how they would approach this assignment including how this would be integrated into the overall work of the evaluation. It is expected that the proposed method would be finalised with the selected contractor through the process of finalising the approach paper. Bidders are expected to identify any areas where they believe data may be insufficient or too limited for the required purposes.

F. Deliverables

The contractor will provide the following deliverables:

- A final VFM approach paper (September 2014)
- Assessment of VFM of the RBA pilot up to end of Y2 (December 2014)
- Assessment of VFM of the RBA pilot up to end of Y3 (July 2015)
- Final report on value (or otherwise) of different approaches to assessing VFM of RBA in education sector (July 2015)

G. Communication

Given the innovative nature of this work, it is expected that there will be considerable interest in this work both in Rwanda and in DFID more broadly. Bidders are requested to explain how they would approach this issue. It is expected that DFID will agree the approach with the selected contractor during the process of agreeing the final approach paper. Current thinking is that there might be at least two learning events

related to this value for money work – one in Rwanda (perhaps to coincide with a Reference Group meeting) and another in DFID offices in either London or East Kilbride.

H. Timing

The consultancy is expected to run from July 2014 to July 2015. Bidders are invited to propose appropriate levels of effort.

I. Expertise required

The bidder is expected to outline how the additional work would be handled within the current evaluation team, explaining implications on level of effort for existing team members and identifying if additional team members are required.

It is expected that this work will require the following skills and expertise:

- A proven track record in the field of development economics (essential) with experience of having published relevant research in peer-reviewed journals (desirable)
- Experience of working with DFID, particularly in relation to assessing value for money of programmes (essential)
- Strong understanding of the conceptual basis of payment by results as implemented by DFID (essential)
- Experience of the education sector in Rwanda (essential)
- Experience of the RBA pilot in education in Rwanda (desirable)
- Good written and verbal communication skills including communicating economic concepts and findings to non-economists (essential)

J. Recipient

The contract supports the effective evaluation of RBA overall. The primary recipient of the output is DFID Rwanda. There are a number of secondary recipients of this work, including DFID's Evaluation, Aid Effectiveness and Value for Money Department.

K. Commercial Proposal

The bidder should outline fully what existing resources will be available from within the current evaluation team for this work (i.e. which personnel and how many days). A clear proposal should be made for what additional resources would be required.

Key documents

Clist, P. (2014) *Approach Paper on Value for Money in Results Based Aid, Rwanda*

Upper Quartile (UQ) and Institute of Policy Action and Research - Rwanda (IPAR) (2014) *Evaluation of Results Based Aid in Rwandan Education – 2013 Evaluation Report*

DFID (2013) *Education Position Paper Improving Learning, Expanding Opportunities* see https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/225715/Education_Position_Paper_July_2013.pdf

Appendix 3 – 2015 Research Approach Paper

Introduction and background

Upper Quartile (UQ) has been commissioned to undertake a mixed-methods process and impact evaluation of the Results Based Aid (RBA) in Education pilot in Rwandan education (2012-2014), considering if, how and in what circumstances the RBA pilot contributed to the results envisaged in the Memorandum of Understanding (MOU) agreed between DFID and the GOR.

The overarching purpose of the evaluation is to determine any **contribution of the RBA pilot to additional learners completing key stages in primary and secondary education**[58] and **additional teachers becoming competent in the use of English as the medium of instruction.**[59]

The evaluation will also consider the response of the recipient (GOR) and other key actors to RBA; the influence of the various factors that impact on achievement of the agreed results; and identify 'lessons learned' to improve the RBA pilot in Rwanda, about the effectiveness of RBA more generally and about how RBA could be transferred to other contexts.

The year 1 report was finalised in March 2013 and the year 2 report in May 2014. This document presents the proposed research and analysis approach for year 3, the final year of the current evaluation, considering specifically the focus, scope and scale of the qualitative research components and the approach to ensuring coherence and complimentarily across the research strands.

This document has been produced in response to comments on the year 2 research from the DFID Specialist Evaluation and Quality Assurance Services (SEQAS) reviewer, regarding a). the appropriate evaluation framework, b). the integration of qualitative and quantitative research components and c). the desire to fill knowledge gaps in the final year of the evaluation. The major issue for discussion and agreement is the appropriate framework to guide the year 3 research, analysis and reporting; other issues largely flow from, and will be addressed upon, agreement of this.

The evaluation framework

While the original terms of reference (TOR) for the evaluation suggested that the appropriate evaluation framework should be the theory of change (TOC) underpinning DFID's wider Rwanda Education Sector Programme (RESP)[60], this was not considered appropriate by the evaluation team. The main reasons are that that the RESP is not specific to RBA; indeed there is no RBA-specific TOC and it was felt that the principle of *recipient discretion,* a defining feature of DFID's approach to RBA, was at odds with a theory-based evaluation approach.

In the Inception Phase the evaluation team agreed with DFID-Rwanda that the evaluation questions (EQs)[61] would be the basis of the evaluation framework. The year 1 research, analysis and reporting was completed on this basis.

In year 2, in response to comments on the year 1 report, the evaluation team worked with a deduced and 'catch-all' TOC for RBA in Rwandan education. This simple TOC, which upholds the principle of recipient discretion, informed the research and analysis process in year 2 (see Upper Quartile, 2015: 7). While the EQs provided the overall structure, the TOC was also included as it was deemed to be of use to DFID in the UK.

Given the changing demands of DFID for evidence on RBA, the complexity of RBA as a modality, and the subtleties emerging in the research findings for Rwanda, it is now considered that neither the original

[58] 'Completion' is defined in the RBA agreement in terms of additional learners sitting key stage examinations in the sixth grade of primary school (P6) and the third and sixth grades of secondary school (S3 and S6).

[59] 'Competency' has been defined in terms of additional teachers reaching level B1 proficiency in relation to the Common European Framework for Reference (CEFR) scale. A baseline sample survey of teachers' proficiency was undertaken by the British Council in 2012 with a follow-up sample survey administered in November/ December 2014.

[60] The results based aid pilot forms part of DFID's £74.98 million GBP Rwanda Education Sector Programme (RESP).

[61] The broad question set was intended to address issues of interest to both DFID-Rwanda and DFID's central Payment By Results (PbR) function.

EQs nor the deduced TOC are suited to providing the nuanced evidence sought. Consequently, DFID has requested consideration be given to a revised evaluation framework for year 3 of the evaluation. The evaluation team met on 3 June 2015 to discuss the options for a revised evaluation framework. The output of that meeting was a draft 'road map' for the year 3 research to be tabled with DFID. The proposed 'road map' is summarised in this document.

Options for a revised evaluation framework

OPTION 1: Evaluation framework based on the status quo

The simplest option for the evaluation team would be to maintain the status quo, with an evaluation framework based on the EQs as they currently stand. This has the benefits of consistency; 3 years of evidence will be easily collated to report findings for the evaluation period as a whole.

However, in maintaining this approach it seems unlikely that the year 3 evaluation will generate substantial 'new' evidence and that the evidence gaps and unanswered questions from year 2 may well persist. This is unsatisfactory to both DFID and the evaluation team.

OPTION 2: Evaluation framework based on a new TOC

DFID (the SEQAS reviewer and Reference Group) appears in favour of the evaluation team developing a bespoke theory of change for RBA in Rwanda. This TOC would be based on the evaluation evidence generated to date, which could be further tested and refined in year 3. While the evaluation team understand DFID's stance on this, we feel it poses a challenge. The major concern is that DFID were clear at the outset that there was no RBA-specific TOC, and the principal of recipient discretion is at odds with an ex ante TOC of specific government actions. Developing an ex post TOC for RBA poses the risk that the existing evidence base is forced into a neat narrative, and is at odds with standard practice of stating which hypotheses will be tested prior to the investigation. It also opens a large space of possible actions that may or may not have been influenced by RBA. Furthermore, the evaluation questions as they currently stand were not grounded in a theory of change. They were developed and agreed on the basis of information needs of DFID-R and DFID-UK. Introducing a TOC as the evaluation framework at this point would create tension between the requirement to test the TOC and to address the EQs.

OPTION 3: Evaluation framework based on tightly defined hypotheses to 'unpack' the existing EQs

Option 3 is the current preferred option of the evaluation team. We accept that the research has outgrown the EQs as they currently stand and that revision is necessary in order to provide the evidence base required by DFID to inform future funding decisions on RBA. However, for the reasons stated above we are unconvinced of the efficacy or practicality of attempting to move towards a newly developed TOC to guide the research at this late stage.

This preferred option represents a *middle ground*: retaining the existing EQs for consistency and practicality of considering the available evaluation evidence base in the round, whilst using the evidence generated to date to more tightly define the research avenues pursued in the hope of filling the remaining evidence gaps and unanswered questions which are of particular interest to DFID. Our proposed approach is presented in section 4 below.

Research approach for preferred option (OPTION 3)

The combined evaluation evidence base to date (including the evaluation in year 1, the evaluation in year 2 and the econometric analysis in year 3) highlights three specific issues which are of interest to DFID and which require further 'unpacking' through qualitative research in the final year of the evaluation. These are:

1. **The significant increase in S3 completion in 2013 and any contribution of RBA to it;**[62]
2. **The significant increase in S6 completion in 2014 and any contribution of RBA to it;**[63]

[62] This is reported in the soon to be published year 2 evaluation report (Upper Quartile, 2015).
[63] This is reported in emerging findings from the year 3 econometric analysis and has been discussed with DFID prior to developing this paper.

3. The contribution of RBA in increasing 'attention' on indicators of English language proficiency and results of this.[64]

The proposed approach in year 3 centres on the **development and testing of research hypotheses around these emerging themes.** This will be complimented with a wider research programme (in keeping with the approach in previous years) to ensure that all the EQs are addressed.

Three research themes

Theme 1: The significant increase in S3 completion in 2013 and any contribution of RBA to it.

Null hypothesis (H0)	RBA did not contribute to the observed increase in S3 completion in 2013
Alt hypothesis (H1)	RBA did contribute to the observed increase in S3 completion in 2013

Research theme 1 represents a gap in the current evidence base that requires further investigation. Econometric analysis observed a statistically significant increase in completion at this grade; that is an increase over and above what would have been expected in the absence of RBA (caveats and limitations of the model withstanding (see Upper Quartile, 2015)).

It was further noted that only 242 students in all of Rwanda are counted as having enrolled in S3 but did not sit the S3 exam.[65] This suggests a concerted effort by Government of Rwanda (i.e. MINEDUC/ REB) and/ or the Districts and schools to achieve completion. Given that S3 is the grade attracting the highest RBA tariff, that RBA contributed to the increase is a potential conclusion.

However, qualitative research in year 2 could not identify any evidence that RBA specifically contributed to the increase in completion at S3, nor did it offer another substantiated explanation for the increase.[66] It is therefore important that this issue is revisited in year 3 in an attempt to respond to the questions raised by the quantitative data (i.e. the increase in S3 completion in 2013 and was RBA a contributing factor). The research approach will explore the null hypothesis that - **RBA did not contribute to the observed increase in S3 completion in 2013 –** using a variety of approaches to triangulate the findings and reach an acceptable evidence based conclusion.

The research will be guided by the existing evidence base, exploring a range of key factors that may explain the S3 increase in 2013, but which could not be either substantiated or rejected by the year 2 evaluation. Relevant factors that may explain the increase in completion (and which have emerged through the evaluation to date) for Research Theme 1 are detailed in **Table A3.1.**

The expectation is that this focused approach to exploring the null hypothesis will allow the evaluation in year 3 to more concretely answer the EQs '*has the RBA approach contributed to impact in relation to the envisaged results?*' and '*what factors have impacted on the achievement of the RBA results?*'[67].

In saying this, it is noted that the main aim of the evaluation, as set out in the TOR is to determine "*whether or not the additional incentive of results-based aid had any effect on the number of children completing different levels of education when compared with what would have been achieved without the provision of this results-based aid*" (DFID, 2012:2). In the absence of identifying a contribution of RBA, the evaluation will seek to identify how any observed increase came about (exploring this is explicit within the year 3 approach), but it is not within the scope of this research to robustly test and rule in or out all possible alternative explanations; there may remain some unanswered questions.[68]

---

64 This is reported in the soon to be published year 2 evaluation report (Upper Quartile, 2015).

65 This can be compared to the preceding year where 5,789 S3 students enrolled but did not take the exam, or expressed as an increase from 93.3% to 99.7% of enrolees completing (Upper Quartile, 2015: 22)

66 Additional exploratory research with the education Senior Management Team (SMT) within Government of Rwanda could not shed light on what was responsible for the remarkable increase in completion at S3 in 2013. This additional research was completed while the year 2 report was being finalised.

67 Note: In relation to this second EQ, the term 'RBA results' refers to completion at S3. As such the evaluation team will be looking to identify the factor(s) contributing to this increase, before considering the extent to which RBA may or may not have been a contributing factor. This approach responds directly to paragraph 16 of the TOR (DFID, 2012: 3).

68 It is noted that paragraph 16 of the TOR refers to determining **'the main drivers'** of additional results. The evaluation team has agreed with DFID the main potential drivers to be the focus of the research in year 3 (i.e. those emerging through the evaluation

to date). The evaluation team will of course be open to other possibilities and unanticipated drivers but for resource reasons it is necessary to target the research effort. This was agreed with DFID in the management meeting on 23 June 2015.

Table A3.1: Approach to Research Theme 1

Theme	Research avenue *'Completion gains due to ...'*	Rationale for this research avenue	Research methods	Contribution to EQs
Research theme 1	Cohort effect	The limitations of the econometric model in controlling for different cohort sizes have been clearly stated throughout. The possibility that the observed effect is due to these difficulties cannot be ruled out and will be explored again in year 3. Some statistical investigations were undertaken in year 2, but the presence of another year of data means cohort size effects can be re-examined with greater accuracy.	▪ **Quantitative analysis**. While not part of the modelling process, statistical techniques can be used to answer and investigate specific questions. For example, the size of different effects in explaining the larger cohort can be calculated, as in last years report. ▪ **Policy and literature review**. Rapid assessment to update years 1 & 2 (this will include search and review of media sources for any leads to explain the S3 increase); ▪ **Review of evaluation notes from years 1 & 2** with a focus on factors relevant to Research Theme 1; ▪ **Qualitative interviews with national level KIs** (see appendix 1); ▪ **Qualitative interviews with District Education Officers** (DEOs) in 3 Districts; ▪ **Qualitative interviews with Sector Education Officers** (SEOs) in 3 Districts; ▪ **Focus Group Discussion (FGD) with school Principals** in 3 Districts.[69] Qualitative interviews and focus group discussions will begin with open questions regarding the S3 increase in 2013 (in particular the remarkable increase in the number of enrolees completing) before probing the specific research avenues identified. Qualitative topic guides will be submitted in advance to DFID for comment and sign-off	▪ **Has the RBA approach contributed to impact in relation to the envisaged results?** **Additional sub-questions:** - Did RBA contribute to the observed increase in S3 completion in 2013? ▪ **What factors have impacted on the achievement of the RBA results?** **Additional sub-questions:** - If RBA is considered to have contributed to the observed increase in S3 completion in 2013, in what ways did it contribute? - If RBA is not considered to have contributed to the observed increase in S3 completion in 2013, what are other possible explanations?
	GOR policy (specifically the impact of 12 YBE)	The announcement of 12YBE in 2011 (initiated in 2012) may have led more students to enrol in S1 in 2011, as they saw a more attractive progression route to the end of secondary school.		
	District level programming, practice and messaging	Qualitative research in year 2 highlighted programming, practice and messaging at District level that District/ Sector and School level KIs consistently linked to the drive from the 'centre' to increase completion. It was further noted that, while these approaches may not be new, there was often a perception of *renewed focus*. Specific approaches include: ▪ **The shift to automatic progression;** ▪ **The School Feeding Programme;** ▪ **Cross community approaches to retaining children in school.** The research in year 2 left unanswered questions around these approaches. While the research in year 2 did not explicitly link these approaches to RBA, we cannot rule out a 'contribution of RBA'. These questions will be posed to national level KIs again in year 3.	
	TVET	In year 2 a senior representative of MINEDUC raised the possibility that enhanced options for progression to TVET may have encouraged completion at S3. While a credible possibility, the evidence in year 2 was inconclusive. This avenue will be explored again in year 3.		

[69] The evaluation in year 3 was never intended to include District or school level research. As such, a pragmatic approach in line with available resources is required. Given that 2yrs has elapsed since the 2013 exams, the qualitative research may be affected by poor recall. We feel the proposed approach is best suited to overcoming this challenge within resource constraints. DEOs/ SEOs have a vested interest in achieving GOR directives; these staff should recall specific actions taken (even given the time lapse) to achieve such high levels of completion. We feel it is also important to gain a school level perspective, by facilitating focus group discussions with Principals we hope to stimulate discussion and unlock further detail that has not emerged in previous individual KI interviews with school level leaders.

Theme 2: The significant increase in S6 completion in 2014 and any contribution of RBA to it.

Null hypothesis (H0)	RBA did not contribute to the observed increase in S6 completion in 2014
Alt hypothesis (H1)	RBA did contribute to the observed increase in S6 completion in 2014

The econometric analysis in year 3 has identified a significant increase in relation to S6 completion. At this stage it appears likely that this 'bump' in S6 results is a cohort effect. Indeed this was predicted by the year 2 evaluation.

> *"This cohort reflects strategic policy changes in Rwanda with the introduction of 9YBE in 2009. It would be expected that this would lead to a larger S1 cohort in 2009, filtering through to a large S3 cohort in 2011 and resulting in a large S6 dividend in 2014; the final year of the current RBA agreement."* (Upper Quartile, 2015: 28)

It seems likely that the 'bump' is related to the introductions of 9YBE and 12YBE, as the years coincide with these significant policy changes. Other potentially interesting avenues to explore in relation to the S6 'bump' are around TVET/ TTC. For students who have decided that an academic track is not for them, the creation of vocational paths which equate to standard S6 exams mean a wider variety of students are likely to have taken this option.

As with Research Theme 1 a mixed-method approach combining quantitative analysis of existing available data with qualitative primary research (national level key informants (KIs), District level staff and school Principals) is proposed (see footnote [10] for further detail). The approach to Research Theme 2 is summarised in **Table A3.2.**

Table A3.2: Approach to Research Theme 2

Theme	Relevant avenue *'Completion gains due to ...'*	Rationale for this research avenue	Research methods	Contribution to EQs
Research theme 2	Cohort effect	The introduction of 9YBE and 12YBE coincide with the larger cohort completing S6 in 2014.	▪ **Quantitative analysis.** The cohort can be traced through the system, and their progression levels can be compared with other cohorts. ▪ **Qualitative research in line with that proposed for Research Theme 1** will also be conducted for research theme 2.	▪ **Has the RBA approach contributed to impact in relation to the envisaged results?** **Additional sub-questions:** - Did RBA contribute to the observed increase in S6 completion in 2014? ▪ **What factors have impacted on the achievement of the RBA results?** **Additional sub-questions:** - If RBA is considered to have contributed to the observed increase in S6 completion in 2014, in what ways did it contribute? - If RBA is not considered to have contributed to the observed increase in S6 completion in 2014, what are other possible explanations?
	GOR policy implementation (specifically impact of 9YBE/ 12 YBE)	The significant reduction in costs, and related awareness of secondary education, is bound to have a positive effect on enrolment.		
	TVET / TTC	The increased focus and availability of vocational training avenues may have encouraged more children to remain in school and complete S6.		

Research theme 3: The contribution of RBA in increasing 'attention' on indicators of English language proficiency and results of this.

Null hypothesis (H0)	RBA did not contribute to the observed increase in teachers' proficiency in English language
Alt hypothesis (H1)	RBA did contribute to the observed increase teachers' proficiency in English language

The follow up survey of teachers' proficiency in English language suggests that gains have been made over the period of the RBA agreement. Further, evaluation evidence in year 2 suggested that the attention given to English language proficiency has intensified over the course of the RBA pilot, with increased focus on the need for policy action. The baseline survey of English language proficiency, a requirement of the RBA MOU, was reported by some to have highlighted major deficiencies in proficiency and reinforced GORs focus; essentially increasing 'attention' in this area.

In the Perakis and Savedoff (2015) view of RBA, the main mechanism for change is through *attention and information*. It is therefore of value for the evaluation in year 3 to further explore this mechanism in the case of Rwanda. The evaluation in year 3 will explore 'attention' in relation to RBA in two ways:

1. The contribution of any observed/ perceived increase in 'attention' to the RBA payment achieved by GOR in May 2015;
2. The contribution of any observed/ perceived increase in 'attention' to GORs ongoing and future approach to English as the Medium of Instruction (EMI).[70]

The above will be articulated as sub-questions of the overarching EQs (see Table 4). This research theme will be explored primarily through key informant interviews with national level stakeholders; both GOR and wider education sector stakeholders (see appendix 1 for an indicative list of KIs).

Unpacking and addressing the EQs

Section 4.1 addresses the key research themes that will be a focus in year 3. This section incorporates these themes into the wider evaluation framework to show our proposed approach to addressing the EQs in their entirety and the way in which the research methods complement one another.

Overall, the core research methods remain unchanged from previous years (the main difference being that the qualitative research has, as planned, been scaled back and will not involve focus group discussions with teachers, parents or pupils). Table 3 summarises the core research methods to be employed in year 3.[71]

A summary of the proposed evaluation framework is shown in Table 4. In response to comments and ongoing discussion with DFID and the SEQAS reviewer, we have revised/ added some additional EQs with the aim of further 'unpacking' the evaluation findings in year 3.[72]

In keeping with the report structure used in year 2 of the evaluation, the evaluation framework is presented in terms of: impact-related questions; process-related questions and additional evaluation questions.

[70] The emerging evidence on this, to be verified and further explored, is presented in the year 2 evaluation report (Upper Quartile, 2015).
[71] Blue text denotes a difference from previous years.
[72] Blue text denotes a difference from previous years.

Table A3.3: Method summary			
Method and research lead	**Description**	**Strengths**	**Weaknesses**
Econometric modelling **(P. Clist)**	Modelling exercise drawing on national level secondary data to identify any effect of RBA over and above what may be expected in its absence. In year 3 qualitative research components have been held back to allow the modelling to be completed before finalising the qualitative research approach. This will allow the qualitative research in year 3 to explore emerging areas of interest in relation to 2014 completion data (specifically the 'bump' in S6 completion – Research Theme 2). Furthermore, statistical analysis will allow various hypotheses to be tested against the data that we have, for example by examining specific drop-out rates.	• Possible to identify SIGNIFICANT change. • Objectivity. • Reliability.	• Cannot answer the 'how' and 'why' questions. • Cannot control for all confounding factors.
Value for money (VFM) analysis **(J. Holden)**	Considers cost-effectiveness of RBA relative to not providing RBA using national level secondary data and standard VfM practice to construct two counterfactuals. The VFM analysis in year 3 will incorporate changes/ additions to the approach agreed following completion of the year 2 research (as discussed at the VfM Round Table meeting with DFID held 3 June 2015). The VfM lead is currently considering options in relation to VFM analysis over the course of the RBA pilot as a whole (2012-2014) and inclusion of below trend results in the VFM model. In addition the VFM model in year 3 will be strengthened with a more complete examination of the assumptions implicit in the model, including the crucial assumptions around the pound for pound equivalence of SBS and RBA aid disbursements for model A, and the assumption around the consistency with averages of the drop-out and progression rates of additional completers in model B.	• Objective basis for considering cost-effectiveness of RBA. • Contribution to the theoretical debate. • Recognised approach allows comparison with other interventions.	• Built on number of contestable assumptions.
Desk based document review **(B. Whitty)**	Desk-based review to situate the RBA pilot within an understanding of the prevailing political and economic processes in Rwanda. In year 3 the review will focus on policy and literature of relevance to Research themes 1, 2 and 3, and on new/ revised policy and literature emerging since completion of the year 2 evaluation. For example, we are aware that RBA funding was not disbursed as SBS in year 3 of the pilot – prior to our KI interviews we will review relevant documents relating to this change. We will also review the evaluation field notes from years 1 and 2 – collating the existing evidence in relation to Research themes 1, 2 and 3.	• Combined with KIIs to explore incentives, relationships, distribution & contestation of power. • Considers how RBA may/ may not function as an incentive.	• Desk based analysis may fail to uncover subtle and ongoing change that is not formally documented.
National level KIIs **(B.Whitty)**	Semi-structured interviews with national level stakeholders to study the three core research themes (the impact evaluation) as well as the response of GOR to the RBA agreement more widely (the process evaluation). An initial, tentative, list of KIs is proposed in Appendix 1. This will be discussed, refined and agreed with DFID prior to commencement of the year 3 qualitative research.	• Insight into stakeholder understanding of RBA and reasoning around incentives.	• Subjectivity. • Researcher presence may influence findings.
District & school-based KIIs & FGDs **(B.Whitty and IPAR-Rwanda)**	Semi-structured interviews with District and Sector education officials and FGDs with Principals in 3 Districts (locations and selection criteria to be discussed and agreed with DFID). The focus of this qualitative research will be in relation to Research Themes 1 and 2.	• Possible to uncover complexities of what is happening where, why, how and with what effect.	• Subjectivity. • Context dependent. • Small sample limits 'generalisability'. • Researcher presence may influence findings.

Table A3.4: Evaluation Framework	Evaluation method					
Evaluation question	Econometric	VFM analysis	Document review	National level	District & school level KIIs/	**Approach and analysis**
Impact-related EQs						
1. What has been achieved?	✓			✓		• Collated summary of completion change and related RBA payments over the course of the pilot (2012-2014). This will draw on the verification reports of HEART, 2013; HEART, 2014; HEART 2015. • Summary of change in English language proficiency and related RBA payments. This will draw on analysis of the British Council English language proficiency baseline survey (Simpson, 2013) and follow up survey [British Council, 2015]. • Key Informant Interviews with British Council lead researchers/ authors of the '2014 Endline Assessment of English Language Proficiency of School Teachers in Rwanda' [British Council, 2015]
2. Has RBA contributed to impact in relation to the envisaged results?	✓		✓	✓	✓	• Collated evidence base presented in relation to Q2a, Q2b and 2c.
a. Did RBA contribute to the observed increase in S3 completion in 2013?	✓		✓	✓	✓	• Econometric modelling and additional quantitative analysis to explore H0 (Research Theme 1). • Desk-based document review focusing on new/ emerging documentation to explore H0 (Research Theme 1). • Review of field research notes from years 1 & 2 of the evaluation to collate the evidence base to explore H0 (Research Theme 1). • National level KIIs (see Appendix 1) to explore H0 (Research Theme 1). • District level KIIs and FGDs (see Appendix 1) to explore H0 (Research Theme 1). • Triangulation of evidence base to determine its strength in relation to rejecting H0 and accepting H1 (Research Theme 1)
b. Did RBA contribute to the observed increase in S6 completion in 2014?	✓		✓	✓	✓	• Econometric modelling and additional quantitative analysis to explore H0 (Research Theme 2). • Desk-based document review focusing on new/ emerging documentation to explore H0 (Research Theme 2). • Review of field research notes from years 1 & 2 of the evaluation to collate the evidence base to explore H0 (Research Theme 2). • National level KIIs (see Appendix 1) to explore H0 (Research Theme 2). • District level KIIs and FGDs (see Appendix 1) to explore H0 (Research Theme 2). • Triangulation of evidence base to determine its strength in relation to rejecting H0 and accepting H1 (Research Theme 2)
c. Did RBA contribute to the observed increase in teachers' proficiency in English that resulted in an RBA payment in 2015?			✓	✓	✓	• Desk-based policy and literature review focusing on new/ emerging documentation to explore H0 (Research Theme 3). • Review of field research notes from years 1 & 2 of the evaluation to collate the evidence base to explore H0 (Research Theme 3).

						• National level KIIs (see Appendix 1) to explore H0 (Research Theme 3). • Triangulation of evidence base to determine its strength in relation to rejecting H0 and accepting H1 (Research Theme 3)
3. What factors have impacted on the achievement of RBA results?			✓	✓	✓	• Collated evidence base presented in relation to Q3a, Q3b and 3c.
a. [If RBA is considered to have contributed to the observed increase in S3 completion in 2013] in what ways did RBA contribute? b. *[If RBA is not considered to have contributed to the observed increase in S3 completion in 2013] what are other possible explanations? [NB: Caveat at section 4.1 applies]*			✓	✓	✓	• Desk-based policy and literature review focusing on new/ emerging documentation in relation to Research Theme 1. • Review of field research notes from years 1 & 2 of the evaluation to collate the evidence base in relation to research Theme 1. • National level KIIs (see Appendix 1) with a focus on Research Theme 1. • District level KIIs and FGDs (see Appendix 1) with a focus on Research Theme 1.
c. [If RBA is considered to have contributed to the observed increase in S6 completion in 2014] in what ways did RBA contribute? *d. [If RBA is not considered to have contributed to the observed increase in S6 completion in 2014] what are other possible explanations? [NB: Caveat at section 4.1 applies]*			✓	✓	✓	• Desk-based policy and literature review focusing on new/ emerging documentation in relation to Research Theme 2. • Review of field research notes from years 1 & 2 of the evaluation to collate the evidence base in relation to research Theme 2. • National level KIIs (see Appendix 1) with a focus on Research Theme 2. • District level KIIs and FGDs (see Appendix 1) with a focus on Research Theme 2.
e. If RBA is considered to have contributed to the observed increase in teachers' proficiency in English language in what ways did RBA contribute? *f. If RBA is not considered to have contributed to the observed increase in teachers' proficiency in English language what are other possible explanations? [NB: Caveat at section 4.1 applies]* *g. To what extent and in what ways has RBA contributed to GORs future approach to EMI?*			✓	✓	✓	• Desk-based policy and literature review focusing on new/ emerging documentation in relation to Research Theme 3. • Review of field research notes from years 1 & 2 of the evaluation to collate the evidence base in relation to research Theme 3. • National level KIIs (see Appendix 1) with a focus on Research Theme 3.
Process related EQs						
4. How is the RBA approach perceived in Rwandan education?			✓	✓		• Desk-based policy and literature review focusing on any new/ emerging documentation in relation to RBA as an aid modality in Rwanda. • Review of field research notes from years 1 & 2 of the evaluation to collate the evidence base in relation to how RBA is perceived in Rwanda? • National level KIIs (see Appendix 1) with a focus on how RBA is perceived in Rwanda? • Triangulation of collated evidence considered against four theories of change for RBA posed by Perkais and Savedoff (2015).
5. How did GOR respond to RBA? *a) Has RBA contributed to programme implementation approaches (such as intensification of implementation)*			✓	✓		• Desk-based policy and literature review focusing on any new/ emerging documentation in relation to GOR's response to RBA.

that complement or enhance sector-wide government plans? *b) Has RBA contributed to programme implementation approaches that detract from sector-wide government plans?* *c) Has the organisational culture of key players such as MINEDUC, REB changed since the introduction of RBA? (For example, has RBA promoted a more results-based approach by government?)* *d) In the context of the relationship between GOR and DFID in Rwanda, would it have been sufficient for DFID to suggest that the envisaged RBA results were a high priority; what, if anything, was the added value of the RBA incentive?* *e) What, if anything, has been the impact of RBA on GOR budgeting and planning and why is this the case?*						• Review of field research notes from years 1 & 2 of the evaluation to collate the evidence base in relation to GOR's response to RBA. • National level KIIs (see Appendix 1) with a focus on the *response* to RBA. • Triangulation of collated evidence considered against four theories of change for RBA posed by Perkais and Savedoff (2015).
Additional EQs						
6. Has Value for Money been achieved? *a) Has value for money been achieved in terms of cost-effectiveness?* b) *Were RBA incentives set at an appropriate level?*	✓	✓		✓		• Value for Money workshop session with DFID [held in London 3 June 2015] • Value for Money analysis (building on the model developed in year 2 of the evaluation) to inform the evaluation's overarching assessment of VfM in relation to RBA in Rwandan education (2012-2014) and to provide a framework for the assessment of VfM of RBA for wider application. • National level KI interviews (including MINEDUC, MINECOFIN and DFID) to consider the implication of different levels of incentive. NB: This assessment will necessarily be qualitative and will represent the perspective of KIs; there is no means to quantitatively test the effect of different incentives.
7. **What lessons have been learned to inform RBA in Rwanda and elsewhere?** *a) Is the RBA approach appropriate in Rwandan education?* *b) What were the circumstances that made the RBA approach appropriate or inappropriate?* *c) What lessons can be learned to support any future RBA negotiations in Rwanda?* d) *If RBA is renewed in Rwanda or introduced elsewhere, what lessons can be learned in terms of how to design and implement the approach?*	✓	✓	✓	✓	✓	• Collated evidence base for Q1-Q6 with a focus on lessons for any subsequent roll out of RBA in Rwandan education; within GOR more widely or in other country contexts.

Indicative implementation workplan

[Removed from published appendix]

Agreements and priority actions

[Removed from published appendix]

List of those to be consulted

Table A3.5: Indicative list of key informants[73]

National level interviews	Org. /Dept	Relevant EQs
Government of Rwanda		
Permanent Secretary	MINEDUC	2,3,4,5,6,7
Director General of Planning	MINEDUC	2,3,4,5,6,7
Head of Department, Primary Education and the Lower Secondary Education	MINEDUC	2,3,4,5,7
Head of Department, Higher Secondary and TVET	MINEDUC	2,3,4,5,7
EMIS Department	MINEDUC	2,3,7
Director General	REB	2,3,4,5,7
Former Director General	REB	2,3,4,5,7
Deputy Director General for Curricula and Materials Production and Distribution Department	REB	2,3,4,5,7
Deputy Director General for Teacher Education Management and Professionalisation	REB	2,3,4,5,7
Head of Education Quality and Standard	REB	2,3,4,5,7
Director, Planning and Research Unit	REB	2,3,4,5,7
Officials responsible for collection and use of data on English completion and sitting exams	REB	2,3,7
External Resources Mobilisation Expert	MINECOFIN	5,6,7
Director General External Financing	MINECOFIN	5,6,7
Officer, External Financing, DFID portfolio	MINECOFIN	5,6,7
Donors		
Education Adviser [Gemma Wilson-Clark]	DFID-R	2,3,4,5,6,7
Independent Consultant involved in design of RBA pilots [Roger Drew]	[on behalf of] DFID-R	2,3,4,5,6,7
Innovative Aid Instruments Adviser [Donald Menzies]	DFID-UK	6,7
Director, Office Education [Susan Bruckner]	USAID-R	2,3,7
Non-Governmental Organisations (NGOs) and other Key Informants		
Lead researchers/ authors of the 2014 Endline Assessment of English Language Proficiency of School Teachers in Rwanda	British Council	2,3,7
Independent consultant involved in scoping/ research in support of the new English as the Medium of Instruction strategy	[on behalf of] DFID-R/ GOR	2,3,7
Other relevant NGOs as identified through snowball sampling (e.g. Girl Hub, Innovation for Education, Plan-Rwanda etc)		2,3,4,5,7
District level interviews		
Key Informants	**Method and sample**	**Number of consultees**
District Education Officers	One Key Informant Interview in each of 3 sample Districts	3
Sector Education Officers	Key Informant Interviews with SEOs in each of the 3 sample Districts	T.B.C
School Principals	One Focus Group Discussion in each of the 3 sample Districts	T.B.C

[73] NB: This is an indicative list of KIs to be discussed, agreed and prioritised with DFID-R. The evaluation team will make every reasonable attempt, within time and budget limitations, to engage priority KIs in the research process and may request DFIDs help to broker key meetings. This was discussed and agreed with DFID in the management call on 26 June 2015. The evaluation team will keep the client informed of any difficulties in reaching priority KIs and the potential implications of this for the research in order that mitigating actions can be taken.

References

British Council (2015) 2014 Endline Assessment of English Language Proficiency of School Teachers in Rwanda; Revised Draft Final Report. British Council for DFID [unpublished].

Perakis, R., and Savedoff, W. (2015) Does Results-Based Aid Change Anything? Pecuniary Interests, Attention, Accountability and Discretion in Four Case Studies: Center for Global Development (CGD) Policy Paper 052, February 2015. Available at: http://www.cgdev.org/sites/default/files/CGD-Policy-Paper-52-Perakis-Savedoff-Does-Results-Based-Aid-Change-Anything.pdf

Simpson, J. (2013) Baseline Assessment of English Language Proficiency of school teachers in Rwanda. British Council Rwanda for DFID [unpublished].

Upper Quartile (2014) Evaluation of Results Based Aid in Rwandan Education: 2013 Evaluation Report. Available at: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/312006/Rwanda-education-results-based-aid-evaluation.pdf

Upper Quartile (2015) Evaluation of Results Based Aid in Rwandan Education: 2013 Evaluation Report [currently unpublished].

Appendix 4 – List of those consulted

Table A4.1: National Level Key Informants (year three)

Organisation	Role	Rationale for inclusion
DFID	[Former] Education Adviser	Responsible Officer, direct experience
DFID	Deputy Programme Manager	Responsible Officer, direct experience
Independent	English Proficiency Consultant	Involved expert in policy context
REB	Director General (DG)	Key Official
Formerly REB	[Former] Director General (DG)	Key Official
REB	Deputy DG Examinations and Accreditation	Key Official
REB	Deputy DG Teacher Management and Professionalism	Key Official
REB	Education Quality and Standards	Key Official
MINECOFIN	External Resources Department	Key Official
MINEDUC	Permanent Secretary	Manage education sector
MINEDUC	Director of Planning	Responsible for planning and budgeting
MINEDUC	Adviser to Minister of State for Primary and Secondary Education	Responsible for priorities in completion
MINEDUC	Adviser to Minister of State for TVET and Higher Education	Responsible for policy on TVET
MINEDUC	Statistician	Responsible for managing data
USAID	Education Adviser	Responsible for determining USAID support to education sector
EPC	Head of Party	Responsible for L3 Programme
Girl Hub	Executive Director	Relevant civil society organisation
PLAN	Adviser	Relevant civil society organisation
UNICEF	Education Adviser	Responsible for UNICEF policy support
IFE	Director	Responsible for management of DFID innovation fund for education
British Council	Multiple roles	Conducted the baseline and endline assessment of language proficiency and delivered a programme work in English proficiency

Table A4.2: National Level Key Informants (year three)

Role	Reason for inclusion	District A	District B	District C
[Formerly titled] District Education Officers (DEOs)	Responsible for implementing education policy/ programming at the district level.	1	1	1
Sector Education Officers (SEOs)	Responsible for implementing education policy/ programming at the sector level.	2	2	2
School Principals	Responsible for district level policy-setting in the education sector.	8	8	7

Appendix 5 – Research tool [National level KIs]

Note to interviewer: Ensure interviewee is provided with an information sheet to introduce the research, outline the role of the consultant and the details of the interview, show how the quantitative work fits into the qualitative work, and ask for permission to take notes, including permission to record.

EQs & Themes	Opening Question	Probing / follow-up questions	Notes	Interviewees
COMPLETION: "I'd like to ask you some questions, first, about completion."				
Research Theme 1; EQ 2 & 3	The quantitative part of the study, which looked at EMIS exam data, shows that S3 completion improved considerably in 2013 from previous years, followed by a drop in 2014. Can you identify reasons for this change?	Probe the answers. If they cite a particular initiative, ask them: (1) when that initiative was introduced; (2) who was it that initiated the initiative, and (3) why these people introduced it. Ask before moving on: "Was there a strong effort to make sure of completion in 2013? Very few did not sit the exam." If so: "Why?" And: "Why did it drop-off afterwards?"	*An open question, so allow to range freely. Provide the relevant explanatory graph and handout of the figures.* *Look out for possible reasons linked to the identified hypotheses.*	PS, MINEDUC; Senior Staff, MINEDUC; Senior Staff, REB; DFID
Research Theme 2; EQ 2 & 3	The same quantitative work shows that S6 completion improved dramatically in 2014, in comparison to previous years. Can you identify reasons for this change?	Probe the answers. If they cite a particular initiative, ask them: (1) when that initiative was introduced; (2) who was it that initiated it; and (3) why they introduced it.	*As above but referencing cohort effects from introducing 9YBE.*	PS MINEDUC; Senior Staff, MINEDUC; Senior Staff, REB; DFID
Research Themes 1 & 2; EQ 2 & 3	Do you think any of the following contributed to either of these changes? • Introduction of 9YBE or 12YBE? • Increased spending in TVET from 2013? • Changes in funding rules or targets for Districts/schools? • Introduction of automatic promotion? • Efforts to work with the community stakeholders? • School feeding?	Can you identify *how* these have helped completion? Is there any way to confirm that, any evidence for that? If the interviewee cites a particular initiative, ask them: (1) when that initiative was introduced; (2) who was it that initiated it; and (3) why they introduced it.	*This is a leading question, so the credibility of the answer lies crucially in how they answer the probing 'how' questions.* *Ask the question by giving them the list in one go, and invite them to home in on specific ones. Then, prompt by 'what about this', 'what about this?' Or 'what about the others?'*	PS MINEDUC; Senior Staff, MINEDUC; Senior Staff, REB
EQ 2 & EQ3	The quantitative work also suggested a reduction in P6 completion in 2014. Can you explain this?		*The limited ability of the system to react to multiple targets may be an explanatory factor for the shifts*	Senior Staff MINEDUC, REB

EQs & Themes	Opening Question	Probing / follow-up questions	Notes	Interviewees
EQ 5	What information does the GoR use in making its decisions and in managing the completion rates? How does it use them?	Has that changed recently? Probe whether they are using the MINEDUC or the REB exams figures. Which of these is considered more robust, more accurate? Are both functional?	*This prepares the ground for exploring shifts in the nature of the attention given to data.* *Visit the office; ask to see the records. Get a sense of how data is collected – for MINEDUC statistics and exam sitting.*	Senior Staff, MINEDUC; REB; EMIS and data collection officials
TEACHERS' PROFICIENCY IN ENGLISH: "I want to turn to Teacher's English Proficiency?"				
Research Theme 3, EQ 2 & 3	A recent study suggests that there have been considerable improvements in teachers' proficiency in English. Can you identify the main reasons for this change?	Probe the answers. Ask: has the mentoring programme been seen as a success? In particular, explore how this has changed in the last year: has there been a recent intensification?	*Adopts the same approach as with completion – beyond the general question, it is worthwhile probing about their perspective on the mentoring programme and whether perspectives have changed recently.*	Senior Staff MINEDUC, REB; other officials; DFID
	Can you explain the changes in the surveys measuring English language proficiency, from the baseline to the endline?	In particular, we are interested in exploring why these changes were made, and who drove them. Do you agree with the test?	*This British Council report was a product of the RBA agreement – discussions prompted by this are therefore prime examples of 'attention'.* *It is important to be clear on the basics of the data collection and use.*	PS, MINEDUC; DG REB; Senior staff REB and MINEDUC; other officials; DFID
	Last year, there were reports that there had been significant discussions around ways forward, following the critical British Council report on English proficiency. Can you provide your understanding of these debates?	Were there any outcomes or changes as a result of the debates? Was there any change in the way the government administered the programme?		
INFLUENCE OF RBA: *"I want to turn to the influence of DFID's influence on implementation."*				
EQ5, Research Theme 1 & 2	Do you think that DFID support to the Rwandan government has had any positive or negative effect on the changes we were talking about earlier, in S3 and S6 completion? If so, how?	Probe for how that support had effects. Refer back to any initiatives mentioned in the previous sections. Remember to probe both for S3 and S6.	*A general question, without specifying RBA at this juncture. Evidence collected thus far suggests that knowledge about the specifics of RBA is limited.*	PS; Senior MINEDUC Staff; DG REB
EQ5, Research Theme 3	Do you think that DFID support to the Rwandan government has had any effect on changes we were talking about in Teacher proficiency? If so, how?	As above.	*As above.*	PS; Senior MINEDUC Staff; DG REB; DFID
EQ 4, EQ5,	Do you think the Results-Based element of DFID's support specifically has made a difference? If so, how?	Probe. If they do not seem to be aware of what we are talking about, ask what they know of the modality.	*Previous experience suggests that many they well not be familiar with this, but it is worth checking.*	PS; Senior MINEDUC Staff; DG REB; DFID

EQs & Themes	Opening Question	Probing / follow-up questions	Notes	Interviewees
EQ6, EQ7	*[For those who indicated some awareness of the modality and the tariffs:]* Do you think the RBA tranche was pitched at the right level?	What level would you suggest? What changes would you make, if DFID were to adopt this modality again?	*Only for those who express awareness of the RBA, likely a small group of very senior officials.*	PS; Senior MINEDUC Staff; DG REB; DFID
EQ5	Can you identify any other changes that the Government of Rwanda has made, in response specifically to the RBA component?	a. Has it helped the government become more results-focused? b. Has it changed how they use statistics? c. Do you think it has changed the organisational culture of MINEDUC or REB, or the Districts? d. Do you think it has helped or hindered the government implement RESP?	*Another leading question; the credibility of the answer is in how they answer the 'how' question* *Seek to tease out RBA-linked targets from GoR / President Kagame results focus etc.*	PS; Senior MINEDUC Staff; DG REB; DFID
Budgetary details for managing RBA tranche				
EQ5	What is the process by which RBA funds are planned for in the annual budget?	How does MINECOFIN enter the RBA amount in its budget planning? Is it £3m (as suggested last year)? If so, why, when that has not happened? Are funds 'earmarked' for MINEDUC? i.e. treated like an EU variable tranche, as suggested last year? What change has there been, recently?	*Last year, we were told RBA funds were earmarked for education, but the education budget was not prepared with reference to particular donor funds, but the overall envelope established by GoR priorities. We understand there have been changes in the way that RBA funds are treated.*	PS MINEDUC; Director, Planning, MINEDUC; DG REB; MINECOFIN
EQ5	If there is a difference between the budgeted amount and the amount DFID disburses, is there any change in-year in the amount that MINEDUC gets?	If there is a shortfall, will the shortfall be borne by MINEDUC, or by the GoR overall?	*This is a key outstanding question from the year two report.*	

Appendix 6 – Research tool [District/ Sector-level officials]

Note to interviewer: Provide an introduction to the study noting that it is an evaluation of DFID's support to the education sector through the Government of Rwanda (but not RBA specifically). Mention that we came to this District the previous year. Ask for permission to record and to write notes, and to use the data.

EQs & Themes	Opening Question	Probing / follow-up questions	Notes	Interviewees
Research Themes 1, increase in S3 sitters in 2013				
Research Theme 1, EQ 2 and 3	The quantitative part of the study, which looked at EMIS exam data, shows that S3 completion improved considerably in 2013 from previous years, followed by a drop in 2014. Can you identify reasons for this change?	Probe any reasons given. If they cite a particular initiative, please ask them to unpack (1) how it achieved its impact; (2) when the particular came into place, and thus how it linked to this particular surge; (3) why there was subsequently a drop again.	*Show the interviewee the handout* *This is the major unexplained element in the evaluation. It is very important to explore this fully, and to probe any answers they may give. Spend time on this, as necessary.*	SEO, DEO
	Very few of those enrolled in S3 in 2013 did not sit the exam in the whole country. Can you explain why the drop-outs were so low in that year?	Probe any reasons given. If they cite a particular initiative, please invite them to unpack it as above.	*As above.*	SEO, DEO
	Do you think that any of the following reasons may explain this: (a) Introduction in 2012 of 12YBE and changes in fees? (b) Management focus on reducing drop-out, through community cooperation etc.? (c) The increase in focus on TVET? (d) The cap on 10% repetition and automatic promotion (e) Changes in the capitation grant (f) Introduction of school-feeding	Please go through each of these potential explanatory factors. Please ask them to unpack (1) how it achieved its impact; (2) when the particular came into place, and thus how it linked to this particular surge; (3)why there was subsequently a drop again.	*These are leading questions. The credibility of the answers lies in the linkage.* *When asking, list all the possible answers, and allow them to identify – but check about the others.* *When probing particular changes, it's important to challenge in particular the link between the initiative & this research theme, that is to say, the S3 increase in 2013. I presume the timing of changes is therefore vital.*	SEO, DEO
Research Theme 2: increase in sitters in S6 in 2014				

EQs & Themes	Opening Question	Probing / follow-up questions	Notes	Interviewees
Research Theme 2, EQ 2 and 3	The same quantitative work shows that S6 completion improved dramatically in 2014, in comparison to previous years. Can you identify reasons for this change?	Probe any reasons given. If they cite a particular initiative, please ask them to unpack (1) how it achieved its impact; (2) when the particular came into place, and thus how it linked to this particular surge; (3) why there was subsequently a drop again.	*The quantitative work suggests that it is a function of a larger group of people entering into secondary school in 2009, due to the introduction of 9YBE. This is a hypothesis that it would be good to test.*	SEO, DEO
	Do you think that any of the following reasons may explain this improvement: (a) Introduction in 2009 of 9YBE and changes in fees, meaning more children in school for that year? (b) Management focus on reducing drop-out, through community cooperation etc.? (c) The increase in focus on TVET? (d) The cap on 10% repetition and automatic promotion (e) Changes in the capitation grant or school funding (f) Introduction of school-feeding	Please go through each of these potential explanatory factors. Please ask them to unpack (1) how it achieved its impact; (2) when the particular came into place, and thus how it linked to this particular surge; (3) why there was subsequently a drop again.	*These are leading questions. The credibility of the answers lies in the linkage.* *When asking, list all the possible answers, and allow them to identify – but check about the others.* *When probing particular changes, it's important to challenge in particular the link with the initiative to this particular research theme, that is to say, the S6 increase in 2014. I presume the timing of changes is therefore vital.*	SEO, DEO
Research Theme 3: increase in teacher proficiency				
Research Theme 3, EQ2 & EQ3	Have you seen a change in the past year in the emphasis central government has placed on teacher proficiency in English?	How has this change been manifested?	*We expect there to have been practical and implementation repercussions of debates following the critical report.*	SEO, DEO
	A recent study suggests that there has been considerable improvements in teachers' proficiency in English. Do you agree with that? Can you identify any reasons for this change?	Probe any reasons given. They are likely to be linked to the Mentoring programme, which was the primary vehicle for addressing this issue. Please ask how this changed.	*This is the equivalent general question to those asked for Research Themes 1 and 2.*	SEO, DEO
P6 completion				
	Lastly, there has been a drop in P6 completion. Can you identify any reasons for this?	Let them answer. Spend a short time on probing this, in particular in relation to a shift away from management prioritisation of this as a focus.	*This may indicate the responsiveness of the system to management attention - as resources and incentives focus on other priorities, past priorities such as primary attendance fall away*	SEO, DEO

Appendix 7 – Focus Group Guide [Principals]

Note to interviewer: Provide an introduction to the study noting that it is an evaluation of DFID's support to the education sector through the Government of Rwanda (but not RBA specifically). Mention that we came to this District the previous year. Ask for permission to record and to write notes, and to use the data.

EQs & Themes	Opening Question	Probing / follow-up questions	Notes
Research Theme 1, EQ 2 & 3	The quantitative part of the study, which looked at EMIS exam data, shows that S3 completion improved considerably in 2013 from previous years, followed by a drop in 2014. Can you identify reasons for this improvement, then drop?	Start with the general question. Then, after they have had the opportunity to respond, please suggest from the list of possible options. As with the previous section, please go through each of the potential explanatory factors that they suggest probing: (1) how it achieved its impact in their school; (2) when the particular initiative came into place, and thus how it linked to this particular surge; (3) why there was subsequently a drop again, and whether that happened in their school.	*This is the major unexplained element in the evaluation. It is very important to explore this fully, and to probe any answers they may give. Spend time on this, as necessary.*
	Do you think that any of the following reasons may explain: (a) Introduction in 2012 of 12YBE and changes in fees? (b) Management focus on reducing drop-out, through community cooperation etc.? (c) The increase in focus on TVET? (d) The cap on repetition and automatic promotion (e) Changes in the capitation grant or school funding (f) Introduction of school-feeding		*As above.*
Research Theme 2, EQ 2 & 3	The numbers on school completion shows that S6 completion improved dramatically in 2014, in comparison to previous years. Can you identify reasons for this change?	Start with the general question. Then, after they have had the opportunity to respond, please suggest from the list of possible options. As with the previous section, please go through each of the potential explanatory factors that they suggest. Please ask them to unpack (1) how it achieved its impact; (2) when the particular came into place, and thus how it linked to this particular surge; (3) why there was subsequently a drop again.	*This question repeats the format of the previous.* *When probing particular changes, it's important to challenge in particular the link with the initiative to this particular research theme, that is to say, the S3 increase in 2013. I presume the timing of changes is therefore vital.*
	Do you think that any of the following reasons may explain this improvement: (a) Introduction in 2009 of 9YBE and changes in fees, meaning more children in school for that year? (b) Management focus on reducing drop-out, through community cooperation etc.? (c) The increase in focus on TVET? (d) The cap on repetition and automatic promotion (e) Changes in the capitation grant or school funding (f) Introduction of school-feeding		
Research Theme 3, EQ 2 and 3	A recent study suggests that there has been considerable improvements in teachers' proficiency in English. Do you agree?		*We suspect that this will be attributed simply to the Mentorship programme, and there will be limited need for follow-up. Less time should be taken for this.*
	Have you seen any particular changes in the past year?		

Appendix 8 – Report of the econometric modelling

[NOTE: This represents the findings of the econometric modelling exercise in isolation. These findings should be read in conjunction with the overarching evaluation report where findings are interpreted in the context of the triangulated evidence base].

List of tables

Table A8.1: Exam sitters by Grade and Year, 2011-2014
Table A8.2: Payment Overview
Table A8.3: Model Summary
Table A8.4: Fixed Effects Regression with Time Trend, 2008-2014
Table A8.5: SUR Estimates with Time Trend, 2008-2014
Table A8.6: P6 Examination sitters estimated with SUR and Panel Data, 2010-2014
Table A8.7: 2012 Payments: Baseline Influence
Table A8.8: Model 2a for Secondary School, 2011-2014
Table A8.9: Model 2b for Secondary level, 2011-2014
Table A8.10: Predicted and Actual District Averages for 2011-2014
Table A8.11: Summary of Results by Approach

List of figures

Figure A8.1: Enrolment, by Year and Grade
Figure A8.2: Changes in Numbers Completing, by District and Exam 2013-2014
Figure A8.3: Out of Sample predictions using Model 2b, by Year, Gender and Level

Data sources

EMIS – The EMIS data available for this year as those to be made available to the general public from the Ministry of Education's website. This includes data on student enrolments, teacher numbers and so forth. The auditor David Dean also made available the examinations data, with permission for DFID-R.

The 2012 General Population and Housing Census provided provisional data for population by district for 2012.

The Integrated Household Living Conditions Survey 3 (EICV3) from 2011 was used to calculate district-level averages that describe important district-level characteristics. These include the average number of repetitions, the fraction of parents reporting no problems at their school and the fraction of the district meeting the definition of "extremely poor".

The Demographic and Health Survey (DHS) 2010 was used in the same way as the EICV3 data. It provides data on those meeting the definition "Poorest", those who "can't read at all" and the fraction of respondents that fall in the 15-19 age range.

Data availability

Data availability is difficult to summarise concisely as the data have not been presented in a consistent fashion and are an amalgamation of several different sources. The largest binding constraint is the set of DEPENDENT VARIABLEs: these are the number of students taking the P6, S3 and S6 examination by district and gender. We have these data for 2008-2014, though some complications creep in for S6 data from 2011 as TTC and TVET examinations were newly administered by the Kigali Institute for Education and the Workforce Development Authority respectively. (This is a problem only indirectly, as students cease to be covered by the EMIS data.)

For enrolment data, crucial for indicating the maximum bound of examination completion, it is available in one of two ways. We use the number of students in a given education level (Primary, Lower Secondary or Upper Secondary) by district, which is available 2010-2014 for Primary and 2011-2014 for Secondary. (The same discrepancy exists for teacher numbers. TVET enrolments were available and so included.) Alternatively, the number of students enrolled by grade is available nationally. This is

used to inform judgements and the general empirical evaluation, but is not included in either Model 1 or 2 as it is not available at the district level.

Many of the district-level characteristics (including population, socio-economic and literacy statistics) are only available at one point of time. This is less than ideal, and its use depends upon the assumption that relative characteristics remain stable. For any characteristics which are likely to change slowly the loss of annual data is negligible.

Glossary

We aim to strike a balance in this report between providing sufficient detail so as to allow an assessment of the quality and breadth of the empirical work undertaken, and providing sufficient clarity so as to allow a non-specialist reader to understand the key messages in the report. There will be inevitable failures, and we are sorry for cases where technical details are glossed over or the reader is left bewildered. In order to aid the non-technical reader a glossary is provided here. Throughout the text, SMALL CAPS are used to denote that a term appears below.

COEFFICIENT– The estimated COEFFICIENT describes the strength of the effect that a one unit increase in the INDEPENDENT VARIABLE has on the dependent variable.

DEPENDENT VARIABLE – In crude terms, the thing we are trying to explain.

DUMMY – A VARIABLE which takes the value one when a condition is met, and zero otherwise. For example, a year DUMMY for 2012 takes the value one when the year is 2012, and zero otherwise. In this example, the COEFFICIENT would measure the effect of it being 2012 relative to the base year.

INDEPENDENT VARIABLE(S) - In crude terms, the thing(s) we are using to explain the dependent variable.

OBSERVATION – One data point. If we have data on the population of each Rwandan district for one year, we have 30 OBSERVATIONS. If we have it for two years, we have 60 OBSERVATIONS.

OUT-OF-SAMPLE Predication – This helps us test the accuracy of our model. We first run the model on a subset of data, deliberately excluding some OBSERVATIONS (specific years or districts). We then use those COEFFICIENTS to 'predict' the excluded OBSERVATIONS. We can then compare the prediction with the known outcome.

SIGNIFICANT – The measure of how likely it is to see an effect purely through chance. To be SIGNIFICANT at the 1% level means that once in 100 times you would see the effect and it would purely be due to chance. The other typical significance levels used are 5% and 10%. The smaller the level of significance, the more confidence the evaluator can have in the evidence.

SPECIFICATION – The list of INDEPENDENT VARIABLES included in a specific model and the type of estimation technique used.

VARIABLE – An indicator or measurement, such as population or teacher numbers.

Introduction

This is the final year of the agreed pilot, and builds upon two previous econometric reports. Much of the modelling and technical work has already happened, and so this report focuses on the results from the model. While this document stands alone as a report, it informs the more general evaluation of RBA in Rwanda. The context, investigation and explanation of results presented here will be in the fuller report.

Background: Payments

Table A8.1: Exam sitters by Grade and Year, 2011-2014

Level	2011	2012	2013	2014
P6	154,954	166,153	163,094	157,123
S3	77,420	80,590	93,732	86,091
S6 Regular	30,878	32,223	30,788	45,973
S6 TVET	15,680	16,669	21,136	20,758
S6 TTC	-	1,410	1,867	2,552

Table A8.1 summarises the number of exam sitters by level and year, with S6 disaggregated into Regular, TVET and TTC. Much of last year's report was devoted to the increase in S3 completers in 2013, with around half of the increase attributed to cohort effects.

This year, the annual changes are negative for P6, S3 and S6 TVET. **There is a *large* jump, however, in those taking the S6 regular exam: an increase of around 50% from a steady base.**

Table A8.2: Payment Overview

Level	2014 - 2013 Improvement	2014 - 2011 Improvement	Total
P6	£51,650	£70,960	£122,610
S3	£0	£90,330	£90,330
S6	£774,600	£235,480	£1,010,080
Total	£826,250	£315,473	£1,223,020

Note: The column 2014-2013 relates to all performance which exceeds the 2011 baseline for the first time, which attracts a tariff of £50 (P6 and S6) or £100 (S3) per student. The 2014-2011 performance relates to increases over the 2011 baseline, but not for the first time, which attracts a tariff of £10 per student.

Table A8.2 shows that in total just over £1.2 million pounds was disbursed for 2014 performance related to improvements in the number of exam sitters. The vast bulk of this was for stronger performance at the S6 level.

Improvements in the level of English for teachers will be discussed in the full evaluation report, but a payment of £1,178,100 was recommended for improvements in English Language skills for 2014. I will now move on to place these improvements in context, using enrolment data which is only available nationally.

Figure A8.1: Enrolment, by Year and Grade

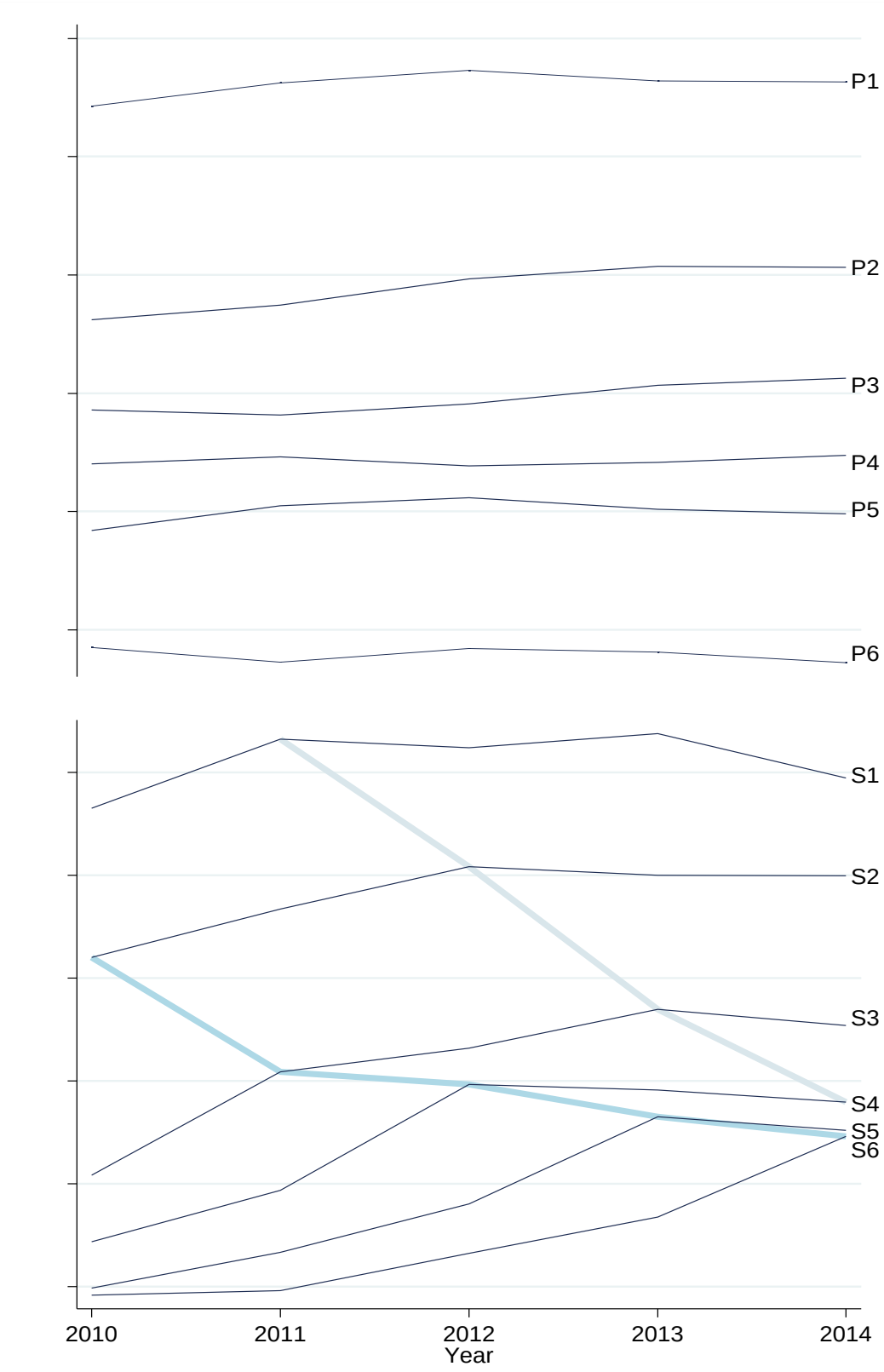


Figure A8.1 shows the number of students enrolled in various grades (not including TVET or TTC, which provides only limited data for per grade enrolment) for the period 2010-2014. The two thick light blue lines trace large cohorts as they progress through secondary school, for example, showing that high S1 enrolment in 2011 filtered through to high S3 enrolment in 2013. Similarly, the high S6 enrolment for 2014 can be traced back to high S2 enrolment in 2010. These two blue bands thus trace large cohorts, which have abnormally large numbers compared to previous and or subsequent years.

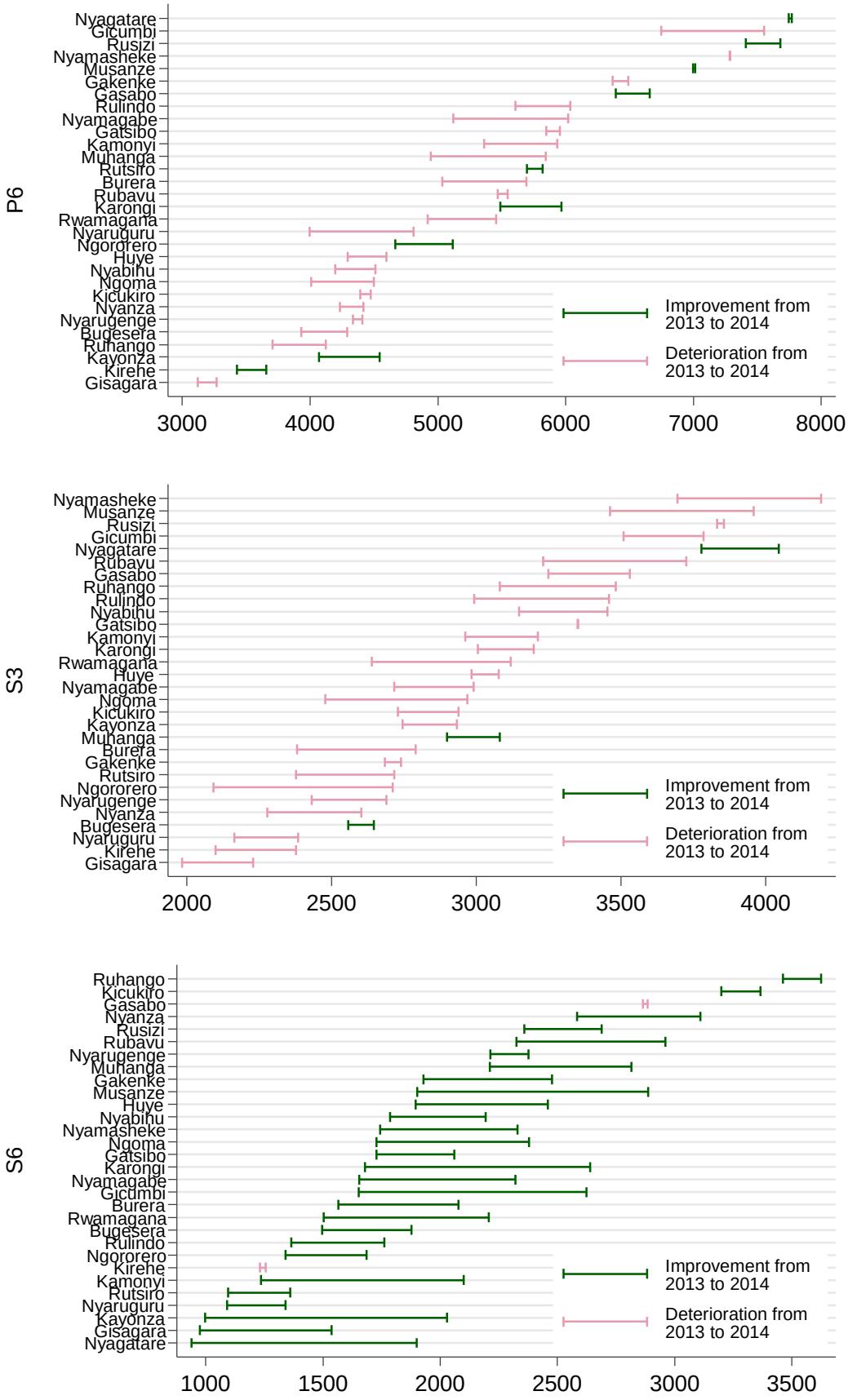
The relatively consistent pattern of enrolments at primary school is clear, with large drop offs in the transition between almost every grade. Very relevant to RBA payments, primary 6 had only 57% of the previous year's P5 enrolments in 2014. This is slightly lower than the 60% average over the period. The figure again points to large repetition rates at the lowest grades, as there are not enough children to be entering P1 for the first time each year: P1 has such large enrolments because of a high number of repeaters. P4, the first grade to be taught in English, consistently has around 85% of the enrolments of P3. This implies that the language change is not as a large a barrier as other factors.

At secondary school level, the highlighted cohorts can be traced through the system. As discussed in last year's report, however, there was a worrying drop off in enrolments for non-incentivised grades. This trend has continued in 2014. A smaller percentage of 2013's P6 enrolees went on to S1 in 2014, and S4 enrolments in 2014 fell despite 2013's S3 enrolments growing. In the context of relatively flat enrolments, the increase in S6 exam sitters is both remarkable and consistent with previous years. It is remarkable because enrolments grew by around 30% in the period 2014-2013. It is consistent because the progression from S5-S6 has been within the range 95-99% for the whole period, with 2014's 95% relatively *low*, if anything. This then appears to be simply a cohort-size effect, at first glance.

Figure A8.2 shows the changes by district and level in exam completion for the last year. At both P6 and S3 there is a general trend of falls in the vast majority of districts. At S3, much of this is 'regression to the mean', where 2013's performance was a relative outlier, and 2014's performance is more in line with expectation. As shown in Table A8.1, the number of students taking the P6 exam has been quite stable over the period 2011-2014, with 2014 performance falling between the high of 2012 and the low of 2011. The largest difference at that level has been the 7% increase over the period 2011-12

At the S6 level, the large improvement has been widespread, with 28 of 30 districts seeing increases (see Figure A8.2). The increase is large: 15,492 more students completed than in 2013. The largest fall in any district for S6 completers (including TVET and TTC) was 23 students (Kirehe), with the largest increase 1,031 (Kayonza). Overall, the picture presented by Figure A8.2 is one of national trends rather than successes at the individual district level. This is not to downplay the effect of district level characteristics (especially in district's relative performance), but rather to emphasise that the time trend is predominately a national effect.

Figure A8.2: Changes in Numbers Completing, by District and Exam 2013-2014



Background: overview of models

As with last year, two main types of model will be used. Model 1 uses no information regarding district characteristics, but instead relies on district-DUMMIES, a time trend and year-DUMMIES. The advantage of the model is that it requires very little data and can thus exploit a longer time series. Model 2 takes a different approach as it uses as much information on district and time differences as possible, including information such as the number of classrooms and teachers in each district. In response to feedback from the reference group[74], Model 2 has been split into two parts. Model 2a uses only information that is available on an annual basis, which allows for district level fixed effects and clustering of the standard errors. Model 2b uses all available information, including data for which we have only one OBSERVATION per district. For example, we do not have annual data on population by district. Model 2a thus ignores differences in population by district, with all static cross-district information being 'soaked up' by the district DUMMIES. Model 2b provides an estimate of the effect of the district-level differences that we do observe. Table A8.3 summarises the model differences.

Table A8.3: Model Summary

	Model 1	Model 2a	Model 2b
Dependent Variable	Exam sitters by district, level and year	Exam sitters by district, level and year	Exam sitters by district, level and year
INDEPENDENT VARIABLES	District and Year DUMMIES, Time trend	All information which is available by district *and* year	All available information
Advantages	Allows standard errors to be clustered and unobserved fixed effects to be controlled for. Allows cross-gender correlation to be controlled for.	Allows standard errors to be clustered and unobserved fixed effects to be controlled for.	Provides an estimate of the effect of all relevant characteristics. Allows cross-gender correlation to be controlled for.
Disadvantages	May 'over fit' the relationship and assumes historical trends continue in perpetuity	Doesn't exploit cross-gender correlation or provide rich policy-relevant information.	May bias standard errors downwards. Assumes homoscedasticity.

Model 1: Results for primary and secondary level

Table A8.4: Fixed Effects Regression with Time Trend, 2008-2014

	P6	S3	S6
VARIABLES	(1)	(2)	(3)
Year	145.5***	248.6***	135.1***
	(4.927)	(15.08)	(4.980)
RBA Year	-136.3*	47.03	-91.98
	(-1.697)	(0.898)	(-1.538)
Constant	-287,365***	-497,605***	-270,048***
	(-4.842)	(-15.03)	(-4.946)
OBSERVATIONS	210	210	210
Number of Districts	30	30	30

[74] In early 2014 the reference group suggested all results used cluster-corrected standard errors. Model 2a is able to do this, but model 2b is not. A large econometric literature exists on the ideal approach with no settled view. We believe the range of options chosen represent best practice.

Note: Cluster-Robust T statistics are provided in parentheses.

Table A8.5: SUR Estimates with Time Trend, 2008-2014

	P6		S3		S6	
VARIABLES	Male	Female	Male	Female	Male	Female
	(1)	(2)	(3)	(4)	(5)	(6)
Year	43.96***	101.5***	109.1***	139.5***	77.07***	57.98***
	(2.819)	(4.928)	(10.90)	(12.93)	(4.170)	(3.183)
RBA Year	-46.71	-89.60	-19.77	66.80	-135.0*	43.01
	(-0.741)	(-1.076)	(-0.489)	(1.533)	(-1.807)	(0.584)
Constant	-86,270***	-201,712***	-218,328***	-279,601***	-154,211***	-116,204***
	(-2.753)	(-4.871)	(-10.85)	(-12.90)	(-4.152)	(-3.174)
OBSERVATIONS	210	210	210	210	210	210
R-squared	0.797	0.819	0.794	0.860	0.609	0.662

Note: Cluster-Robust T statistics are provided in parentheses.

Table A8.4 presents a simple test of RBA by grade, summing any RBA effect across the three years into one estimate per grade.

Table A8.5 allows for two effects per grade, as it disaggregates any effect by gender. In both cases district characteristics are controlled for using district fixed effects. In all columns of both tables, there is no evidence of a SIGNIFICANT positive RBA effect. The general upward trend of completers in Rwanda is found in every case, and any RBA effect would need to be identified over and above these general time trends. In no case is RBA found to be positive and SIGNIFICANT, and in two cases (P6 in Table A8.4 and S6 Males in Table A8.5) it is found to be SIGNIFICANT and negative. Model 2 will examine whether the yearly changes are SIGNIFICANT in themselves, and control for more information.

Model 2: Results for primary level

Table A8.6 presents two sets of results. Columns (1) and (2) are more similar to the model 1 SPECIFICATION, with the additional controls for the numbers of students enrolled and teacher numbers. The effect of RBA is captured through the use of the year DUMMIES for 2012, 2013 and 2014. If there were a positive RBA effect, one would expect to see positive COEFFICIENT estimates for these years. The only SIGNIFICANT year effects are that 2011 and 2014 were SIGNIFICANTLY worse years than the general trend.

Columns (3) and (4) come from a SUR regression, where errors in the numbers of completers by district are allowed to be correlated across the two genders. The use of SUR is advantageous in many ways, as it allows an estimate of the effect of components which we do not have annual data for. For example, we only have one estimate of population for each district, and the SUR model allows this effect size to be estimated. The results of columns (1) and (2) are echoed in columns (3) and (4): RBA years, where SIGNIFICANTLY different from trend, were worse than the general trend. This model also emphasises that the baseline year (2011) was SIGNIFICANTLY below-trend, which will have led to higher RBA payments than would have otherwise been the case. In other words, had 2010 been the baseline, the RBA payments for P6 performance would have been much lower. This would have been most relevant in 2012, where £616,550 was disbursed for P6 performance. This appears to have been partly due to below-trend performance in 2011: this is investigated further below.

Table A8.6: P6 Examination sitters estimated with SUR and Panel Data, 2010-2014

Model:	2a: Fixed Effects		2b: SUR	
	Male	Female	Male	Female
	(1)	(2)	(3)	(4)
(Fe)Male Primary School Students Enrolled	0.047*	0.037	0.044***	0.074***
	(1.71)	(1.16)	(6.71)	(7.31)
Male Primary School Teachers	-0.082	-0.043	0.64***	0.53
	(0.35)	(0.13)	(2.59)	(1.40)
Female Primary School Teachers	-0.077	0.18	0.41*	1.39***
	(0.38)	(0.67)	(1.72)	(3.91)
% of district deemed “Poorest” (DHS)			153.1	274.8
			(0.62)	(0.74)
Average Repetition, district (EICV)			-1329.5***	-1790.1***
			(4.88)	(4.42)
% of district reporting “No School Problems”			162.2	1287.4**
			(0.46)	(2.41)
% of District that Can’t read at all			-2919.2***	-4275.5***
			(6.18)	(5.94)
(Fe)Male Population by district			0.0040***	-0.0017
			(3.88)	(0.93)
2011	-132.6***	-121.1**	-215.6***	-261.8**
	(3.76)	(2.37)	(3.07)	(2.46)
2012	-41.0	69.2	-76.8	-52.3
	(0.65)	(0.86)	(1.14)	(0.51)
2013	-90.1	7.45	-121.2*	-119.3
	(1.38)	(0.09)	(1.80)	(1.17)
2014	-177.7**	-99.7	-225.1***	-252.0**
	(2.54)	(1.10)	(3.32)	(2.46)
Constant	787.1	1426.9	2121.1***	1845.1**
	(0.79)	(1.22)	(4.19)	(2.42)
OBSERVATIONS	150	150	150	150
R Squared (overall)	0.37	0.41	0.74	0.68
Chi Squared			475.89	355.49

Note: Robust T stats are given in parentheses. SUR T-stats are not cluster-robust, FE are.

Table A8.7 illustrates the effect that the abnormally low P6 performance in 2011 had on the amount transferred, by imagining that 2010 had instead been the baseline for P6. The transfer would have been lower by £229,250: a 40% decrease for P6 performance. Of course, a different year may have meant higher disbursements for other grades. The point of this exercise is to note that normal fluctuations in the baseline year can have large effects on specific payments.

Table A8.7: 2012 Payments: Baseline Influence

Level	Payment using Actual 2011 Baseline	Payment using Hypothetical 2010 Baseline	Implied Drop in Transfer
P6 Male	£216,450	£104,750	£111,700
P6 Female	£346,500	£228,950	£117,550
Total	£562,950	£333,700	£229,250

Summary: Primary level

For primary school completion, there is no evidence of any above-trend performance in any of the years in which RBA was active. Furthermore, models 1, 2a and 2b each report some performance which is SIGNIFICANTLY *below* trend. Model 2b provides the most negative estimates, reporting 2014 completion was an average of 477 students lower than trend per district: 14,310 students nationally. While model 1 is likely to overestimate possible progress given previous improvements, other models are not prone to this weakness. Indeed, Figure A8.1 shows large drop offs in enrolment year to year. To put the below trend performance into perspective, 14,310 extra students completing primary education could be achieved by having an extra 5% of P5 enrolees take the P6 exam a year later, from a base of only 61%.

To summarise the evidence for P6, all models agree that the years in which RBA was active were either on or below trend. Performance in the baseline year of 2011 was also poor, which may have influenced the perception of subsequent years.

Model 2: Secondary level

Table A8.8: Model 2a for Secondary School, 2011-2014

	S3		S6	
	Male	Female	Male	Female
	(1)	(2)	(3)	(4)
Lagged P6/S3 Completers	-0.041	-0.033	-0.036	0.30**
	(0.55)	(0.65)	(0.17)	(2.16)
Lower/Upper Secondary Enrolment	0.036**	0.035*	0.024	0.030**
	(2.16)	(2.01)	(1.28)	(2.23)
Male Secondary Teachers	0.82	0.48	1.31*	1.46***
	(1.03)	(0.76)	(1.81)	(2.79)
Female Secondary Teachers	-0.020	0.098	-0.73	-0.13
	(0.01)	(0.07)	(0.38)	(0.08)
2012 Dummy	-70.2	74.6	-145.1**	-42.4
	(1.36)	(1.69)	(2.18)	(0.74)
2013 Dummy	98.0	262.2***	188.3	167.0
	(1.13)	(3.44)	(1.58)	(1.51)
2014 Dummy	-44.0	141.8	435.3**	250.5
	(0.34)	(1.16)	(2.39)	(1.38)
Constant	613.8	794.1**	17.3	-509.0
	(1.60)	(2.40)	(0.04)	(1.43)
OBSERVATIONS	120	120	120	120
R Squared (Overall)	0.60	0.46	0.66	0.73

Note: The results here were estimated with district fixed effects, and cluster-robust t statistics are reported in parentheses.

Moving to secondary results Table A8.8 presents results for Model 2a, which allows for district level fixed effects but is only able to control for a limited number of district-level factors. The controls that are included all have COEFFICIENTS of the expected sign, where SIGNIFICANT. Specifically, keeping other factors equal an increase in enrolment at the lower or upper secondary school level of thirty students increases the expected number of completers by around one, for both genders and levels. The lagged number of completers is only SIGNIFICANT for S6 females, and means that for each additional S3 completer we should expect 0.3 extra S6 completers three years later. The number of teachers is generally INSIGNIFICANT, apart from male teachers in S6. Unfortunately, data is not available on the level teachers are (i.e. we cannot distinguish between upper and lower secondary school teachers) but the result implies that there are important differences in teacher numbers by district. Looking into the EMIS numbers more closely, the aggregate *pupil:teacher* ratio at (standard) secondary level ranges from 30.3

(Nyagatare) to 20.2 (Gakenke). (As noted last year, the INSIGNIFICANT effects of female teachers are likely due to their low numbers at secondary school.) To illustrate the magnitude of these estimates, imagine Nyagatere added 65 male teachers so that it's *pupil:teacher* ratio was equal to Gakenke's. This implies an extra 85 students completing S3 and 180 students completing S6, year on year (this effect would presumably be through lower class sizes). The financial benefit of doing so under RBA would be £17,500 in the first year.

On RBA effects, Table A8.8 gives mixed results. It shows two positive results (S3 females in 2013 and S6 males in 2014) and one negative result (S6 males in 2012). Table A9.9, which is able to control for more district-level difference, paints a similar picture but with more SIGNIFICANT results. Table A9.9 reports three negative effects in RBA years, and five positive effects. Last year's report dealt with the 2013 increase in S3 completion, finding that much of the increase could simply be attributed to cohort effects which the model was not able to adequately control for (due to the paucity of pre-RBA data).

Figure A8.1 is useful in this respect, as it puts the cohort-size effects into context. The positive effect at S6 in 2014 will be investigated further in the full report.

On non-year DUMMIES, Table A8.8 generally conforms to expectations. Districts will tend to have more completions if they have:

- Higher lagged completion. An additional 4-5 students completing P6 in 2011 is consistent with an extra S3 completer in 2014.
- A lower percentage of the district in the lowest poverty category. Moving from the district with the highest incidence of people in the poorest category to that with the lowest is consistent with an extra 102 S3 male completers, 79 S3 female completers, 257 S6 male completers and 230 S6 female completers.
- Higher enrolment at that stage. For both genders and levels, an extra 20 enrolled students at that stage (e.g. across lower secondary school) is consistent with around one extra completer, even after controlling for lagged completion.
- More teachers. Again, there is an effect such that students benefit more from teachers of their own gender, with seven of the eight effects positive. However, female teachers appear to have a *negative* effect for S3 boys. This is most likely simply controlling for the number of male teachers i.e. that an extra female teacher will tend to mean one less male teacher, other things being constant.

As with last year's report, the COEFFICIENTS on population are somewhat counterintuitive, as they imply higher population leads to lower completion. However, the effect is conditional on enrolment and lagged completion, and so should be interpreted as saying that districts which do poorly in terms of enrolment and P6/S6 completion *for their population* also tend to see lower levels of completion at S3/S6 *for their population*.

Table A9.9: Model 2b for Secondary level, 2011-2014

	S3		S6	
	Male	Female	Male	Female
(Fe)Male Population, 2012 Provisional Data	0.00017	-0.0015**	-0.00052	-0.00079
	(0.30)	(2.40)	(0.73)	(1.26)
Lagged P6/S3 Completers	0.19***	0.23***	0.042	0.15*
	(4.63)	(7.50)	(0.46)	(1.92)
% of district deemed "Poorest" (DHS)	-206.9*	-161.3	-523.1***	-468.7***
	(1.80)	(1.31)	(3.16)	(3.49)
Lower/Upper Secondary Enrolment	0.047***	0.050***	0.067***	0.055***
	(6.17)	(5.61)	(4.93)	(4.78)
Male Secondary Teachers	0.97***	0.56***	1.07***	0.90***
	(5.06)	(2.60)	(3.25)	(3.39)
Female Secondary Teachers	-0.90**	1.08**	0.97*	2.17***
	(2.26)	(2.17)	(1.79)	(4.28)
2012 Dummy	-70.9**	25.7	-199.8***	-73.8*
	(2.00)	(0.63)	(3.82)	(1.68)
2013 Dummy	133.0***	208.1***	64.8	109.1**
	(3.38)	(4.61)	(1.16)	(2.31)
2014 Dummy	-21.2	7.65	252.1***	200.7***
	(0.51)	(0.16)	(3.94)	(3.69)
Yrs15_19			-567.0	-395.9
			(0.80)	(0.66)
Constant	-7.57	10.3	-68.8	-261.1
	(0.08)	(0.09)	(0.32)	(1.41)
OBSERVATIONS	120	120	120	120
R Squared	0.73	0.78	0.76	0.82
Chi Squared	343.17	474.65	382.23	572.95

Note: These are Seemingly Unrelated Regressions with T statistics provided in parentheses.

Summary: Lower secondary (S3)

For two of the three years in which RBA operated, the evidence is of either INSIGNIFICANT or SIGNIFICANTLY negative year effects. For the remaining year (2013) there is a SIGNIFICANTLY positive effect, which is found by Model 2a and 2b, most strongly for girls. This year's report has added an extra year of data, which has modestly increased the ability of the model to control for cohort-size effects (as it allows another view of the extent to which lagged completion of P6 affects S3 completion). This has led to smaller COEFFICIENT estimates for both models and genders for 2013's S3 performance, but these are mostly quite small and remain SIGNIFICANT.

To be specific, last year's results from Model 2a estimated that 2013 performance was 12,810 higher than trend. This year's Model 2a estimates that performance was only 10,800 higher: the 2013 effect has been revised downwards by 2,010. Model 2b's estimate has only reduced by 360, to 10,230. These two Models are now much closer to each other, and allow the evidence to be reassessed in light of the new data.

Summary: Upper secondary (S6)

Model 1 shows that taken as a whole RBA years have been below trend. Model 2a and Model 2b disaggregate the effects by year, and paint a mixed picture with 2012 SIGNIFICANTLY below trend and 2014 SIGNIFICANTLY above trend. In Models 2a and 2b the magnitude of the positive effect is greater than the negative effect.

The simple summary statistics are useful in providing the stark facts: 2014 saw S6 exam completion jump by around 50%. While the 2012 negative effect may lead the 2014 positive effect to be overstated, it is clear that the positive effect should be investigated properly. Initial judgements imply that the effect is likely to be due to the large cohort, but this will be explored in the fuller report.

Out of Sample predictions

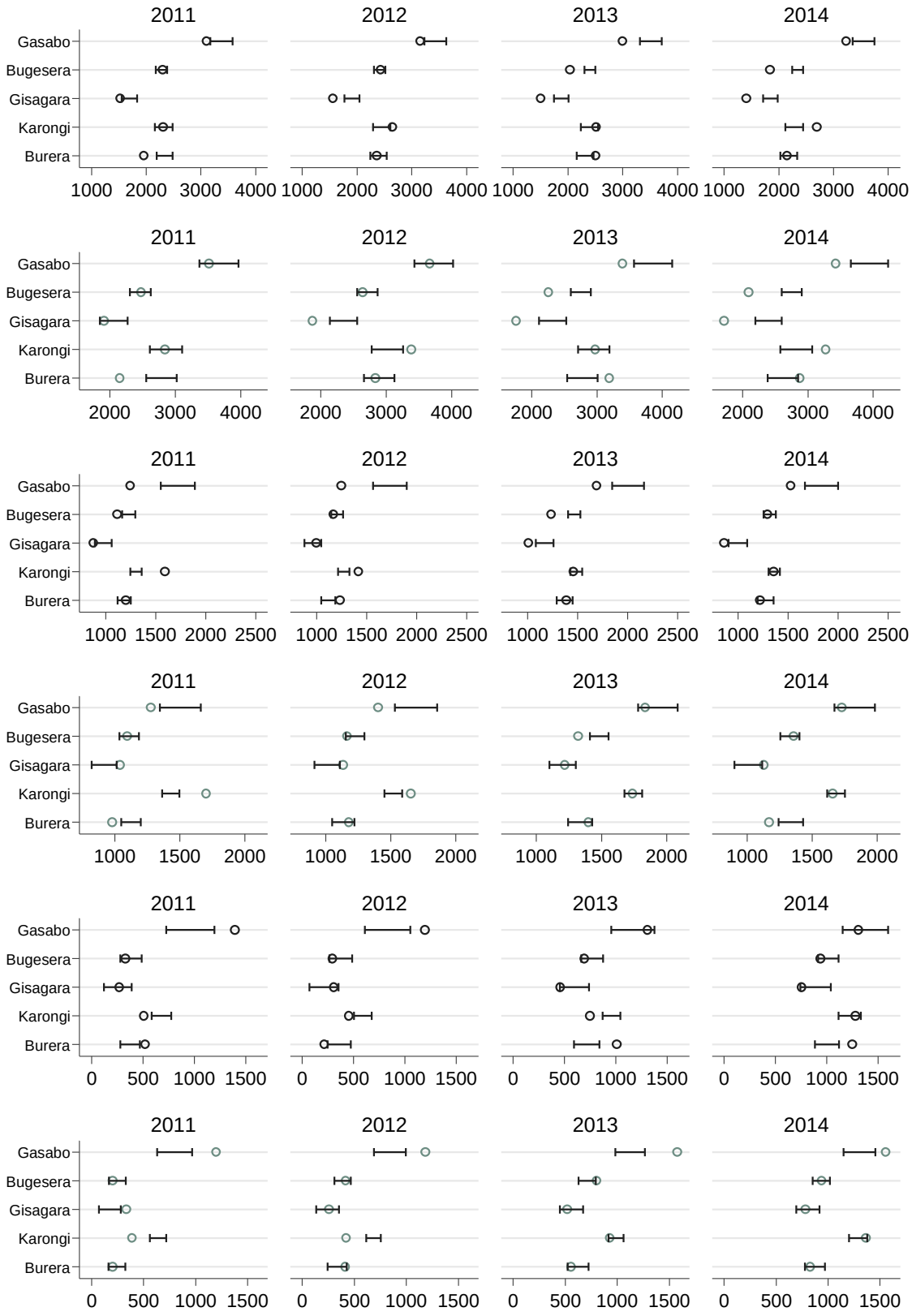
As with previous years, OUT-OF-SAMPLE predictions are used to gauge the accuracy of the model. The specific model used for these predictions are from the SUR regressions (model 2B): they should not be taken as a judgement on the accuracy of other models. Five districts were randomly chosen (in 2013) to be 'out of sample', with the model run for the remaining 25. The COEFFICIENTS were then used to predict OUT-OF-SAMPLE performance for the five districts for the years 2011-2014. Figure A8.3 presents the 95% confidence interval that the model predicts as a line, and the actual performance as a circle. If the model is performing well, the circle should be within that line.

To summarise the evidence, correlations between point predictions (not the range graphed) and actual results by grade and gender for 2014 range from a low of 0.73 to a high of 0.92, with an average of 0.85. Given the data limitations and that these are made completely out of sample (i.e. the model does not know of these other 5 regions), these predictions are surprisingly good. It should be remembered that even a perfect model would see actual performance outside of the 95% confidence interval 5% of the time. The consistent weakness of the model is that it finds Gasabo boys difficult to predict at all levels, but for S6 boys the model has performed well for 2013 and 2014. One reassuring feature of the errors in Figure A8.3 is that errors are made both by predicting too high and by predicting too low. Table A8.10 summarises these errors, by summing the average error across the five OUT-OF-SAMPLE districts and four years. Given that the test is stringent (as it removes one sixth of the data), the model performs well with errors of 1.4-7.8%.

Table 13: Predicted and Actual District Averages for 2011-2014

	Predicted District Average	Actual District Average	Average Error
P6 Male	2463.8	2311.6	152.2
P6 Female	2885.7	2712.6	173.1
S3 Male	1349.5	1256.5	93
S3 Female	1377.8	1359.1	18.7
S6 Male	736.6	761.6	-25
S6 Female	690.0	743.5	-53.5

Figure A8.3: Out of Sample predictions using Model 2b, by Year, Gender and Level



Note: circles represent the actual results, with the lines representing the 95% out of sample prediction.

Conclusion

The econometric report uses a variety of approaches, allowing the synthesis of each modelling approach to bring greater certainty to the findings. These insights are summarised by model and level in Table A8.11. It is clear that RBA had no SIGNIFICANT effect on performance in the majority of cases, including primary school in all years. In several cases a SIGNIFICANT negative effect is found. We have not given this effect much credence, as they have typically been on specific years with limited corroboration from other models and little theoretical basis for further exploration. There is possibly a question mark over whether the fall in primary school completion has been related to RBA, or more generally the switch to focus on other areas (e.g. completion). Two specific incidents are SIGNIFICANTLY positive, and corroborated by several models: 2013 S3 performance and 2014 S6 performance. The underlying reasons for both of these will be investigated further in the main report.

Table A8.11: Summary of Results by Approach

Level:	P6	S3	S6
Summary Statistics	Unambiguously, 2014 has been a poor year for P6 completion.	2014 was much worse than 2013 for S3 completion, but this appears to be related to 2013 being an exceptionally good year partly due to cohort-size effects.	2014 saw a remarkable 50% increase in Regular S6 completion. This appears to be a cohort-size effect.
Model 1	RBA was negative where SIGNIFICANT, and never positive.	RBA was INSIGNIFICANT, but possibly better for girls.	RBA was negative where SIGNIFICANT, and possibly better for girls.
Model 2a	The baseline was SIGNIFICANTLY below trend. Even against this, RBA was occasionally SIGNIFICANT and negative (2014 for boys)	RBA-years were generally on trend. 2013 was a SIGNIFICANTLY good year for girls, but this could have been a cohort-size effect.	RBA-years were generally on trend, but for boys 2012 was SIGNIFICANTLY low and 2014 SIGNIFICANTLY high. This could be a cohort size effect.
Model 2b	The baseline was SIGNIFICANTLY below trend. Even against this, RBA was occasionally SIGNIFICANT and negative (2014 for boys and girls, 2013 for boys).	In RBA-years, 2013 was SIGNIFICANTLY above trend for both genders and 2012 SIGNIFICANTLY below for boys. Cohort effects are difficult to control for.	Both genders were SIGNIFICANTLY below trend in 2012. In 2013 for girls and 2014 for boys, completion was above trend. Cohort effects are difficult to control for.

Appendix 9 – Detailed report of the Value for Money Assessment

[NOTE: This represents the findings of the econometric modelling exercise in isolation. These findings should be read in conjunction with the overarching evaluation report where findings are interpreted in the context of the triangulated evidence base].

Summary

The degree to which results based aid (RBA) provides value for money (VfM) to DFID and the British taxpayer is a key question for its efficacy as an aid instrument. The RBA agreement was intended to help drive change in the education sector as aligned with priorities of the Government of Rwanda (GoR). The question of whether the RBA pilot provides VfM is therefore one of how much benefit is derived from DFID support to education through money disbursed by RBA, and whether RBA is a better means of delivering that aid than the likes of sector budget support (SBS); put slightly differently - *whether a payment by results (PbR) mechanism delivers through the incentives not present in other, non-PbR forms of aid*.

Our approach was therefore developed for the assessment of the VfM of RBA to education in Rwanda, with a model developed for the 2014 evaluation report. The approach has been continued for the 2015 report with some adaptations and extensions to the model to try to ensure greater realism and rigour of the estimates produced. For the 2015 report, the model is also run for all the three years of the pilot - 2012, 2013 and 2014.

Summary of the approach

The VfM approach as proposed in the original design paper is to look at the cost effectiveness of RBA relative to not providing RBA. There are two components to this, based on two different counterfactuals:

A. The VfM of aid spent on education through the RBA instrument, compared to the counterfactual of not providing that aid to education.

B. The VfM of aid spent on RBA, compared to the counterfactual of still providing that aid to education, but by another instrument without any explicit incentive present – for example by sector budget support (SBS).

C.

The model presents two versions of test B – B1 a 'naïve' estimate using the benefits derived from the number of 'extra sitters' at P6, S3, and S6, which the RBA design pays out upon, i.e. assuming a zero trend. The second, B2, uses the econometric model of the RBA evaluation to estimate the number of 'additional sitters' at these grades – i.e. the *statistically significant* level of change, taking into account changes above or below existing trends – and is therefore the more rigorous of the tests.

Overall VfM findings

The model is now run for each of the three years of the RBA evaluation, and the overall results of this are shown in Table A9.1. The overall finding is that aid to education in general provides good VfM (results for test A), but less dramatically than for the 2014 report with the greater realism added to the model, the benefit-cost ratio (BCR) being 2.6, an estimated £11.2 million in benefits accruing as returns to education from £4.4 million disbursed in aid.

In terms of the test of RBA as an incentive mechanism, the naïve estimate (test B1) still finds significant and large benefits, with a BCR of 2.2 and net present value (NPV) of over £100 million for the project period. However the more plausible test, test B2 has a much smaller result with a NPV of £18 million and a BCR of 1.1 – though it should be noted that this includes costs that do not incur to DFID directly. The main reason for the much narrower result this time is that statistically significant negative changes are added to the model for test B2, with 2012 then generating a negative NPV, only partially offset by 2013 and 2014.

Table A9.1: Summary of results from VfM assessment of RBA for all years (GBP million, in 2015 GBP)[75]

	PV benefits (£ million)	PV costs (£ million)	NPV (£ million)	B-C ratio
Test A				
Benefits attributable to the 2012 year of education	3.0			
RBA disbursed for 2012		1.2		
Benefits attributable to the 2013 year of education	4.9			
RBA disbursed for 2013		1.9		
Benefits attributable to the 2014 year of education	3.2			
RBA disbursed for 2014		1.2		
Overall NPV for test A	**11.2**	**4.4**	**6.8**	**2.6**
Test B1				
Net benefits from extra sitters in 2012	65.6	34.7	30.8	1.9
Net benefits from extra sitters in 2013	74.5	33.4	41.2	2.2
Net benefits from extra sitters in 2014	63.2	24.5	38.7	2.6
Evaluation costs and verification costs		0.7		
Overall NPV for test B1	**203.3**	**93.4**	**110.0**	**2.2**
Test B2				
Net benefits from extra sitters in 2012	21.1	52.1	-31.0	0.4
Net benefits from extra sitters in 2013	72.7	41.0	31.7	1.8
Net benefits from extra sitters in 2014	82.0	64.1	17.9	1.3
Evaluation costs and verification costs		0.7		
Overall NPV for test B2	**175.8**	**157.9**	**17.9**	**1.1**
Tests A and B combined				
Overall A+B1	**214.5**	**97.7**	**116.8**	**2.2**
Overall A+B2	**187.0**	**162.3**	**24.7**	**1.2**

Combining the two models gives an overall assessment of the VfM of RBA, with a benefit-cost ratio of 2.2 for A+B1 and 1.2 for A+B2, and a net present value (NPV) of £117 million and £25 million for the two combinations respectively. Sensitivity analysis is conducted in the paper, finding that NPV returns remain positive even when a number of assumptions are made more stringent.

As before, the results here need to be heavily caveated that they are based on a large number of assumptions, and in addition are reliant on the econometric analysis undertaken for the evaluation with further assumptions therefore implicit. Finally, the test of whether RBA is VfM as an incentive mechanism depends on whether there is evidence to say that it is or not; this evidence does not come from the VfM model, but should come from a rounded view of the quantitative and qualitative evidence of the RBA evaluation on whether this is the case.

Cost effectiveness per sitter - VfM findings

Given all of the assumptions noted, it is instructive to look at the individual VfM for the changes that RBA intends to bring about. Table A9.2 provides the unit breakdowns for test B2 in terms of how much

[75] The benefits and costs from 2012 and 2013 are uprated into 2015 GBP using the UK Retail Price Index (RPI).

benefit is derived from an *additional sitter* at each of the grades, P6, S3 and S6. This is informative more generally in terms of the scale of change that is required to make RBA good VfM. It shows that the NPV per individual additional sitter ranges from around £1,000 to £2,500, with this benefit the additional lifetime earnings as a result of extra years of schooling, offset by the cost of that schooling. This compares well to the tariff of £50 or £100 depending on the grade paid out by the RBA mechanism.

The result of this is that a relatively low number of additional sitters are required to break-even on the additional costs from the RBA mechanism – the evaluation and verification costs. Thus, 493 additional individuals completing P6 would be the break-even point, or 228 at S3, or 221 at S6. This shows that if RBA can be shown to have a significant and positive incentive effect, then even at a relatively small scale it should be VfM from the point of view of the cost-effectiveness model generated for this paper.

Table A9.2: Analysis per additional sitter for test B2

Additional sitter at Grade:	PV benefits	PV costs	NPV per individual additional sitter	Number of sitters required to break-even on additional RBA costs (evaluation + verification)
P6	£2,976	£1,856	£1,120	493
S3	£4,499	£2,079	£2,419	228
S6	£4,080	£1,584	£2,496	221

Introduction

For the RBA pilot to provide VfM to DFID and the British taxpayer it should be shown that the benefits outweigh the costs. This means answering the questions of how much benefit is derived from DFID support to education through the Government of Rwanda (GoR), what the best means of delivering that aid is, and whether the 'payment by results' (PbR) incentive mechanism of RBA delivers more relative to its cost than other non-PbR forms of aid.

An approach was developed for the assessment of the VfM of RBA to education in Rwanda, with a model developed for the 2014 report. The approach is continued for the 2015 report with some adaptations and extensions to the model, to try to ensure greater realism and rigour of the estimates produced. For this report, the model is also run for all the three years of the pilot - 2012, 2013 and 2014.

The approach to measuring VfM

Measuring the VfM of aid to education represents a complex problem. This is because the principal benefits in terms of returns to education come only when the educated are in employment, and then accrue for many years afterwards. This requires estimates of how long each individual stays in education, when individuals will enter and exit the labour market, and what returns in terms of wages will be in future years. Inevitably such a model requires a number of assumptions. Assumptions are stated and justified as far as possible in this paper.

The VfM approach, as proposed in the VfM design paper, is to look at the cost effectiveness of RBA relative to not providing RBA. There are two components to this, based on two different counterfactuals:

A. The VfM of aid spent on education through the RBA instrument, compared to the counterfactual of not providing that aid to education.
B. The VfM of aid spent on RBA, compared to the counterfactual of still providing that aid to education, but by another instrument – for example by sector budget support (SBS).

The overall VfM of RBA will then be A+B: that is, the effectiveness relative to the cost of the aid itself (A); and the effectiveness of RBA relative to providing aid in another form (B). The test of most interest in looking at RBA as opposed to other aid modalities will be B, as this can be taken to be the narrower test of the hypothesis that RBA functions through an incentive effect that is not present in other forms of aid. This test is based on the accompanying econometric evaluation of RBA, with discussion of the

qualitative case also crucial for assessing the strength of evidence for the assumptions the model requires.

The model presents two versions of test B – B1 using the benefits derived from the number of 'extra sitters' at P6, S3, and S6, which the RBA design pays out upon. The second, B2, uses the econometric model of the RBA evaluation to estimate the number of 'additional sitters' at these grades – i.e. the *statistically significant* level of change.

The model presented here is built on existing data of the Rwandan education system, with the key variables being the rate of drop-out, the rate of repetition, and the rate of progression between grades. This is then combined with labour market information on wage rates and returns to additional years of education, to model the returns to education over time, and specifically those returns attributable to the years of investment in education in question, in this case 2012, 2013, and 2014.

Adapting the VfM model for the 2015 report

Partly in response to comments on the 2014 report, and also due to the use and availability of additional data and reflection on this data, the model for the 2015 report has been modified and strengthened with a number of adaptions to ensure a greater degree of realism and robustness. The principle changes that have been made include:

- For the 2015 model, data from the years 2012, 2013 and 2014 are used for estimating drop-out rates, repetition rates and progression rates. The data is used to estimate a range of estimates with a 'low drop-out and repetition rate' scenario using the lowest figures from these years, and a 'high drop-out and repetition rate' scenario using the highest figures. The average of these two figures is then used in the final model.
- Returns to education estimates are refined. This draws on greater analysis of the income levels in Rwanda and more discussion of the accuracy of returns. In addition, regression analysis is conducted using data from the WageIndicator Foundation survey carried out in 2012 (Besamusca et al., 2013) to estimate returns to schooling in Rwanda, drawing on this as additional evidence for the returns estimates used.
- Unemployment and economic inactivity have been added to the model, using information from the Integrated Household Living Conditions Survey 3 (EICV 3) survey, a large household survey conducted across the country between November 2010 and November 2011.
- In addition the initial base level of wages for someone with no education is revised downwards – this is based on a closer look at Rwandan GDP data together with the Gini coefficient, and information from EICV3 on the number of hours worked per week.
- A higher estimate of real wage growth is used, based on data from the EICV3 survey and the earlier EICV2 survey conducted in 2005/2006.
- Mortality rates have been added to add greater realism to the labour market model.
- Four more scenarios are added for sensitivity analysis, including for positive or negative changes to the level of unemployment and economic activity, and two scenarios for the relative value of aid to education compared to other forms of funding.
- For the test B2, the results of statistically significant negative changes in the numbers of sitters at P6, S3 and S6 are modeled as well as positive changes.
- Break-even points are discussed in more detail, including how these compare to the level of tariffs.

Assumptions made in the VfM model

Despite attempting to provide a greater degree of realism, a model such as this projecting so far into the future is inevitably built on a large number of assumptions. These try to be as justified as possible, but they should be clearly stated and interrogated. These include:

- A discount rate of 10 per cent - this is the standard DFID discount rate, including for Rwanda. As the model seeks to measure returns to education in the labour market from those in

education now, the returns are often far into the future, while most of the costs are incurred in the present.

- Drop-out rates, repetition and progression rates all continue into the future. This does not take any account of changes to supply, quality or demand for education in coming years.
- There is no drop in quality that would impinge on the likely returns to education of individuals leaving schooling at various levels- The underlying assumption of relatively constant educational quality was not firmly investigated given the difficulty of doing so.
- Labour demand will continue to be sufficient to absorb increasing amounts of individuals in the market with an average higher level of education with no significant changes to unemployment, economic inactivity or underemployment.
- Much of the assumed benefit in the VfM model is derived from the implied extra years of schooling over and above the incentivised year. If extra schooling for additional sitters was found to be lower than for other students – i.e. an 'additional sitter' at P6, S3 or S6 was subsequently more likely to drop-out than other students - then the size of the benefits would fall substantially.

The essence of the assumptions made in the VfM model is that current trends will continue into the future. In reality, change can be discontinuous. With respect to RBA, it is very important to note the central assumption that RBA itself does not represent a discontinuity – i.e. RBA does not change behaviour around other aspects of the system that would impact on educational outcomes. This is particularly relevant with respect to the quality of education, which is likely to be the main underlying driver of economic and non-economic returns to education.

Changes made to the VfM model for the 2015 report

The model developed for 2014 attempted to provide a realistic framework for both progress through the Rwandan education system, and then experience in the labour market for individuals when they leave education. As the sensitivity analysis for the 2014 report demonstrated, the model was very sensitive to returns to education in particular – unsurprising as this is the determinant of future wages which formed the estimated benefits of the model. As a result of this, greater scrutiny is placed on the returns estimates here, with new estimates used for the 2015 model informed by a closer look at the evidence. In addition, the model attempts to create greater realism in the labour market modelling, including the presence of unemployment and economic inactivity, as well as including mortality rates. The changes are summarised here, with the net effect reducing the estimates from the 2014 model.

Returns to education

The returns to education used in the 2014 model were estimated based on international research by Psacharopoulos and Patrinos (2004), estimates for Africa from Schultz (2004), estimates from Rwanda from Lassibille and Tan (2005), and survey data for Rwanda conducted by the WageIndicator Foundation in 2012 (Besamusca et al. 2013). The final returns used were primary returns of 15 per cent, junior secondary of 15 per cent, upper secondary of 20 per cent, and tertiary returns of 20 per cent.

A closer analysis of more recent international returns data, particularly data collected for the World Bank's 2013 World Development Report, shows that these estimates are potentially on the high side. Montenegro and Patrinos (2012) find that average returns to schooling in SSA are 12.8 per cent, while for low-income countries more generally average returns are 10.5 per cent. Breaking down, SSA average rates are 13.4 per cent for primary, 10.8 per cent for secondary, 21.9 per cent for tertiary. Montenegro and Patrinos also provide evidence that the higher the average years of schooling in a country, the lower the returns, with an additional average year of schooling reducing average returns to education by around 0.4 per cent - *"Higher point estimates of the returns to schooling are associated with lower levels of schooling in a country. This suggests that schooling increases respond to price signals. In other words, it could be that as demand goes up and the supply follows, the price tends to fall."* In short, an expansion of schooling is likely to lead to decreasing rates of return.

Data from the WageIndicator Foundation, which conducted a survey in Rwanda in 2012 was also used. The survey was conducted between late October and early December 2012, with a total of 2,074

persons interviewed in towns in all provinces of Rwanda. The survey found a median net hourly wage of the total sample of 450 Rwandan francs (RWF) (£0.44), with the respondents working an average of 60 hours per week and 5.9 days. The hourly wage for someone with tertiary education was estimated at 1,369 RWF per hour (£1.34). This indicated an average rate of return of 18 per cent for every year of schooling (the average for primary, lower secondary, upper secondary and tertiary).

Regression estimates of returns to education in Rwanda

For the 2015 model and report, to come up with a more rigorous estimate of returns from the WageIndicator data, we ran an econometric regression using the Mincerian returns formula to estimate returns to additional years of schooling in Rwanda. The so-called Mincerian wage regression is named after its instigator, Jacob Mincer, and involves estimating an empirical relationship by regressing the log of earnings on years of schooling:

$$\log y = \log y_0 + rS + \beta_1 X + \beta_2 X^2$$

The dependent variable **y** is earnings (y_0 is the level of earnings of an individual with no education and no experience), with explanatory variables including **S**, the years (or levels) of schooling completed; **X** is years of potential labour market experience, with X^2 indicating the wage increase with work experience but at a decreasing rate (the coefficient β_2 being negative). The coefficient **r** in such an equation is then interpreted as the *rate of return* to an additional year (or level) of schooling in higher lifetime earnings.

An enhanced Mincerian model uses dummies for different levels of education to estimate their individual specific returns – isolating the returns to primary, secondary and tertiary. This is the regression run for this paper using the WageIndicator data, specified as follows:

$$\text{Log } Y = \alpha + \beta_1 EXP + \beta_2 EXPSQ + \beta_3 P + \beta_4 LS + \beta_5 US + \beta_6 NT + \beta_7 T$$

Y = earnings per hour

α = intercept

β_1 = return per year of work experience, **EXP** = years of work experience

β_2 = return on years of experience squared, **EXPSQ** = square of years experience

β_3 = primary school returns estimate, **P** = primary school dummy (as highest level).

β_4 = lower secondary school returns estimate, **LS** = lower secondary dummy

β_5 = upper secondary school returns estimate, **US** = upper-secondary dummy

β_6= post-secondary (non-tertiary) returns estimate, **NT** = post-secondary dummy

β_7 = tertiary level returns estimate, **T** = tertiary dummy for highest level reached.

The final estimate for returns to a year of schooling at each level is calculated by the difference between each progressive level, so for example the upper secondary estimate uses $\beta_5 - \beta_4$, and then the inverse power of the amount of years for that level is used to estimate the returns per year - for upper secondary which is for 3 years, the estimate is calculated by $(1+(\beta_5 - \beta_4))^{(1/3)}$. The model gives the following returns (detailed results in Annex):

> Primary = 1.5%; Lower secondary = 18.7%; Upper secondary = 24.2%; Post-secondary (non-tertiary) = 22.4%; Tertiary = 25.4%.[76]

Issues with the WageIndicator data

The clear anomaly of the regression is the very low returns estimated for primary level. This indicates there may be biases in the data, and indeed a closer look finds such biases. This is connected to a warning about returns estimates discussed in Montenegro and Patrinos (2012): *"For cost or*

[76] The tertiary estimate is based on assuming 4 years at that level, if 3 years is used the return is 36.7 per cent. The post-secondary estimate assumes 3 years at this level. Both levels of secondary are 3 years, and primary is 6 years.

convenience, surveys may concentrate on subpopulations that are easier or less expensive to reach, focusing on firms rather than households or on urban populations while excluding rural residents."

The WageIndicator data reported in Besamusca et al. (2013) contains such a bias due to the sampling strategy used, which did not come from random household sampling, but rather sampling *"based on a country-level sampling frame of establishments, a random sample of the establishments was drafted. From the randomly sampled establishments a list of workers from a broad range of occupations was interviewed."* Surveying from establishments on a government or private sector database is likely be unrepresentative in an economy still dominated by household farming such as Rwanda, and where the vast majority of the workforce works in the informal economy (EICV3). There were some other biases that were clear from the data itself:

- Those in the WageIndicator sample were more educated than the Rwandan population as a whole. In the sample, 28.2 per cent had tertiary education as the highest level compared to 3.5 per cent that hadn't finished primary education. However in reality in Rwanda of those aged 15-24, 58 per cent have not completed primary education, while just 1 per cent have went on to post-secondary education (EPDC, 2014).
- The mean annual salary in the WageIndicator data-set was RWF2.9 million (£2,848), and the median annual salary RWF1.2 million (£1,179). This is significantly higher than the Rwandan average, with the median adult annual income found to be RWF143,000 (£140) (from EICV3 data for 2011)
- The WageIndicator sample overrepresented richer parts of the country, in particular the Northern and Kigali provinces.
- The EICV3 found 3 out of 5 million adults in Rwanda to be working as independent farmers; with a further half a million working as waged farm workers (over 800,000 were then *wage non-farm*, with close to half a million working as *independent non-farm*). However in the WageIndicator survey, just 1.7 per cent of those surveyed worked in agriculture, forestry and fishing, only 13.6 per cent of those sampled lived in a country village, with 65 per cent of the sample coming from a small city (10k-100k people).

So in summary, the WageIndicator data should be treated with caution, as should the returns estimates generated from the data.

Final returns estimate used

The final returns used then takes account of the larger span of estimates reported, as shown in Table 3, including the international averages from Psacharopoulos and Patrinos (2004) and Montenegro and Patrinos (2012), and the Rwandan data from Lassibille and Tan (2005) and the regression run using WageIndicator data from Besamusca et al. (2013). The final estimates then involve changes from the 2014 model estimates, including a reduction from 15 per cent to 10 per cent for returns to primary education, a reduction from 20 per cent to 17.5 per cent for upper secondary, and an increase from 20 per cent to 25 per cent for tertiary level. Estimates are then all slightly lower than the average of returns estimates from the different sources. The returns estimates are further calibrated below.

Table A9.3: Private economic rates of return to education estimates including for 2015 VfM model

Source	Primary	Lower secondary	Upper secondary	Tertiary
Psacharopoulos and Patrinos (2004) for low-income countries	25.8%	19.9%		26%
Montenegro and Patrinos (2012) for SSA	13.4%	10.8%		21.9%
Lassibille and Tan (2005) Rwanda estimate based on	19.4%	29%		33.3%

1999-2001 Household Living Conditions Survey				
Estimates used for 2014 model	15%	15%	20%	20%
Econometric result based on data collected by Besamusca et al. (2013)	1.5%	18.7%	24.2%	25.4%
Final estimates used for 2015 model	**10%**	**15%**	**17.5%**	**25%**

Calibrating returns to primary education

Given the radically different estimate of primary returns from the regression conducted with other estimates, further calibration was carried out for looking at returns to primary education through a mix of education data and the EICV3 survey.

Using the EICV3 estimate of average income per adult or RWF141,000 (£139), and annual consumption growth of 4.4 per cent from 2006 to 2011 measured by EICV3 and EICV2, with inflation of 3.7 per cent in 2011 to 2012 and 6.1 per cent for the following year, gives an estimate of average annual income of RWF169,000 (£166) in 2013. Given that those employed make up 84.2 per cent of the population, the income per employed adult would be higher at RWF200,000 (£196).

If this is the average income per employed adult, then for the returns estimates to be well-calibrated, this should tally with the average number of years of schooling. According to the latest Human Development Report data, the average Rwandan aged 25 years or over has just 3.3 years of education[77]. This then tallies well with the average wage figure for none educated workers in 2013 of RWF 153,000 (£150). The average education figure of 3 years would equate with an approximate 10 per cent annual return for primary education, which is the figure used for the 2015 model.

The base level of wages and real wage growth

Another key component of the model is the initial level of income for someone with no education, which is used as the basis to estimate wages for everyone in the labour market model depending on their number of years of schooling. The estimate used in the 2014 model was RWF306,000 (£301), which was based on the WageIndicator data, where the hourly wage for someone without education was estimated at RWF98 per hour (£0.10), and in turn they were estimated to be working a 60-hour week. For the 2015 model, this number has been looked at and revised. This is in part due to looking at the likely income level in Rwanda when assessing the country's GDP per capita alongside its Gini coefficient (a measure of inequality, where the more unequal a country, the lower the median wage is compared to the average GDP per capita figure). The Gini coefficient in Rwanda was estimated by the EICV3 at 0.49. This means that while the GDP per capita in Rwanda (in 2013) was around £426, this is misleading as an average income per person figure due to the high level of inequality. The EICV3 found that the ratio of the income the 90th percentile to the 10th percentile in Rwanda to be over 6. This means that the poorest 10 per cent of Rwandans would have an income of closer to £100 – compared to an average for the richest 10 per cent of over £750.

In addition, the estimate of an average 60-hour week from the WageIndicator data is questionable. The EICV3 survey found an average working week closer to 30 hours[78]. The 30-hour figure is therefore used for the 2015 model, applying this to the average hourly wage in 2013 of RWF98 (£0.10), and giving an average annual wage for someone without any education of RWF153,000 in 2013 (£150).

[77] Rwanda ranks 168th out of 187 countries on UNDP data on the number of years schooling of those aged 25 years or over, with just 3.3 years of education. - http://hdr.undp.org/en/content/mean-years-schooling-adults-years (up from 2 years in the year 2000. [Data in the tables are those available to the Human Development Report Office as of 15 November, 2013].

[78] EICV3 found the average number of hours worked a week in all jobs is 31 (men 35, women 28) but this falls to 27 (30 men, 25 women) for those whose main occupation is agriculture.

This figure appears to be more closely calibrated to the figure found for income in EICV3, where the average income per adult was RWF143,000 (£140), and a more robust measure of average consumption was found to be RWF124,000 (£122) per adult. Rwandans are therefore much poorer than the GDP per capita figure alone might suggest, which is now reflected in the 2015 model.

In addition to the change to the base level of income, the real wage growth rate is changed for the 2015 model. For the 2014 model the figure used was 2.5 per cent. This time, the model takes into account the difference between EICV3 and EICV2 survey data, for which the more rigorous estimate was consumption, where annual consumption growth was 4.4 per cent from 2006 to 2011. As a result, the real wage growth estimate used for the 2015 model is 4 per cent.

The labour market in Rwanda, unemployment and economic inactivity

The labour market of a poor sub-Saharan African (SSA) country such as Rwanda has a number of factors that should be taken into account to ensure a greater degree of realism, in particular the prevalence of unemployment and economic inactivity as well as underemployment. The majority of the workforce still operates in subsistence farming, and there is debate to whether returns to education estimates hold for this group, and whether the international averages discussed above often derived from the formal sector may not be representative of the average adult. As a result of these debates, the 2015 model adds an assumption that there is a proportion of the labour force economically inactive or unemployed at any given time.

According to the EICV3, economic activity rates in Rwanda are very high. Of the total population aged 16 and above, 84.2 per cent were employed at the time of the 2010/11 survey; just 0.8 per cent were technically unemployed; 3.6 per cent were inactive and not seeking work; and 11.3 per cent were in education. Given that the VfM model includes those in education, it is those that are unemployed or inactive that must be added to the model. Using the EICV3 report, this is therefore estimated at 4.4 per cent of the labour force.

Veracity of returns estimates in a case such as Rwanda

While Kingdon and Söderbom (2008b) in Pakistan find that returns to education in agriculture are similar to those in other occupations, in Ghana Kingdon and Söderbom (2008a) find *"There are no significant returns to literacy or numeracy in agriculture for either age group, suggesting that Ghanaian agriculture is mainly traditional in that cognitive skills that would allow a person to, for example, follow instructions on fertilizer packs does not raise agricultural earnings."* It is not clear where Rwanda would fall, but the Ghanaian evidence should serve as a warning around returns to education in subsistence agriculture.

As Golub and Hayat (2014) argue, in Africa, pervasive underemployment rather than open unemployment may be the norm in agriculture and urban informal sectors. There are two factors that may be at play here, one is low demand for labour generally, and the other that those educated do not have the skills employers require. Indeed, as for many other African countries (Garcia and Fares, 2008), in Rwanda those with more education are more likely to be unemployed. Unemployment among young people is higher in urban areas (9 per cent) than in rural areas (3 per cent), and the highest unemployment rates are observed among young university-educated women (MIFOTRA 2013).

There is some debate over whether education itself leads to higher returns, or whether education is correlated with other factors that lead to those returns – for example those from higher income or social class groups. The standard assumption is that greater education means a higher supply of skills leading to higher productivity, and this productivity is reflected in higher wages or *returns to education.* This is built on perfectly competitive labour markets, where:

> *"If labour markets are perfectly competitive then an employer can find any number of equally productive workers at the prevailing market wage so that a worker who left could be costlessly replaced by an identical worker paid the same wage. And a worker who lost their job could immediately find another identical employer paying the same wage so would not suffer losses."* Manning (2010).

However a number of studies have questioned whether this holds. Mortensen (2005) finds that education explains only about 30 per cent of variations in labour compensation, and labour heterogeneity appears to be robust to the inclusion of numerous controls, where *"large wage differences across employers suggest the existence of significant match rents".* These rents are caused by the fact there are costs for the unemployed to find jobs and costs for employers to find workers, and also costs for job losses to both (Manning, 2010). In addition, workers have imperfect knowledge of wages available elsewhere, and employers have imperfect knowledge of the productivity of workers (due to their inability to monitor at all times, the basis of efficiency wage theories).

The presence of rents may mean that productivity and wages do not match. Indeed, Fox and Oviedo (2008) show that wage returns do not reflect productivity differences in sub-Saharan Africa (SSA). Söderbom et al. (2002) find using data from Ghanaian manufacturing sector, that when observable and unobservable aspects of human capital are controlled for, wages are much higher in larger firms, with human capital of minor importance in explaining either the distribution of earnings or productivity across firms of differing size.

A survey by the Rwandan Private Sector Federation found that when asked to name the main barriers to entrepreneurship, entrepreneurs stated that good communication skills and management skills were the biggest barriers - higher than barriers of risk or collateral for credit (PSF and OTF Group 2008). Furthermore the Rwandan Human Resource Development Agency (HIDA) in a 2008 survey found that large skills gaps across Rwanda's public and private sector were present, with skills not meeting the labour demand at that time (HIDA, 2009). More recent research found that 45 per cent of recent Rwandan Masters graduates were underemployed – raising questions of the qualities of skills graduates had acquired and the ability of the labour market to absorb more educated workers (New Times, 2015).

This questions whether there is a constraint to the demand for skills as well as the supply of skills, which could imply that expanding education could actually reduce average returns. The same study (New Times, 2015) also linked the unavailability of work to lack of connections from influential persons. These findings provide a warning about the assumptions underlying labour market modeling using returns to education estimates. This is reflected in the sensitivity analysis for the 2015 model looking at different returns scenarios, reported below.

Mortality rates

The 2013 model did not take mortality into account. While mortality rates for any given year are very low - from 0.2 per cent of 15 to 19 year olds to 1.2 per cent from 50 to 59 year olds - the compound effects year-on-year imply that the share of a given cohort in the labour force is significantly smaller the more years into the future the model projects. For example of those aged 18 in 2013, three quarters would still be around in 2046 aged 51, and 62 per cent in 2054 when they would be aged 59. Mortality is therefore added into the model this time, with a significant reduction on returns (a reduction of returns of approximately 7 per cent). Mortality data comes from the 2012 census (MINECOFIN and NISR, 2014).

Negative changes as well as positive changes for model B

The 2014 model for test B1 and test B2 looked only at positive changes to the numbers of sitters at P6, S3 and S6. This was justified on the basis that the incentive effect to increase completion was the theory of change being tested, so a null hypothesis of no change would only be rejected if the change was positive. However, comments to the 2014 VfM model as well as greater reflection on the possible meaning of negative results have led to their inclusion for the 2015 model for test B2.

This is a major change to the model and significantly reduces the net benefits generated under the test B2 – which is because for some of the years and grades in question, statistically significant negative results were experienced.

Including the negative results as well as the positive results involves a reversal between the costs and benefits in the model. In the case where the number of additional sitters at P6, S3 or S6 is positive, the benefits are the additional net lifetime earnings gained as a result of the extra years of education these

individuals experience, set against the cost of the extra schooling they accrue. But where the change in completion is negative, the net lifetime earnings will fall, and so is counted as a cost in the model, and this is set against a benefit of the saving from expenditure on education not incurred as a result of earlier drop-out.

TEST A: VfM against counterfactual of amount of aid not allocated to education

The first test of VfM looks at the benefits of aid allocated through RBA against a counterfactual of not providing that aid to education. As such it can be seen as not simply a test of the effectiveness of RBA, but of other forms of aid to education, including aid disbursed as SBS. This method will therefore look at how much aid contributes to the overall benefits attributable to the education system compared to the cost of the aid provided.

Costs under test A

The cost element for test A is quite simply the amount of money disbursed to the Government of Rwanda in RBA. For the case of RBA in Rwanda the direct cost of aid was originally up to £9 million over the full period of the pilot dependent on the results experienced. Disbursements have now been made and form the costs for test A: £1,164,150 paid in 2013 for results achieved in 2012; £1,883,420 paid in 2014 for improvements made in 2013; and £1,223,020 paid in 2015 for improvements made in 2014; a total of around £4.3 million.

The VfM assessment for test A is based on examining this contribution of aid relative to the total of Rwandan expenditure on education. The key assumption here is that each pound spent on education is equal to every other pound, and that the aid provided is non-fungible. The assumption here is that the money contributes towards spending to education for each year to which they were disbursed. The model assumes that this money is additional to GoR expenditure on education.

Breakdown of GoR education expenditure

The total spend on education is required in order to estimate the share of the benefits from this expenditure delivered by aid, and in this case, RBA. In Rwanda, education expenditure is made up of a mixture of internal resources and development aid. UNESCO (2012), estimated the following split for 2009/2010:

- Internally-financed Ministry of Finance budget to education = 46 per cent
- Ministry of Finance from General Budget Support (GBS) aid = 25 per cent
- Sector Budget Support (SBS) = 18 per cent
- On-budget project aid = 1 per cent
- Off-budget project aid = 10 per cent

More recent data from the Rwanda 2013/2014 budget used here for the VfM model shows that the contribution of GBS and SBS aid has fallen since 2009/2010. Table A9.4 shows the breakdown between GoR Ministry expenditure (MINEDUC) - making up around 60 per cent of Government spend, principally for higher education and technical and vocational education and training (TVET), and districts - making up 40 per cent of expenditure and mainly for pre-primary, primary and secondary education.

Table A9.4: Breakdown of GoR expenditure on education in 2013/2014

Budget line	Budget to	Amount RWF (billion)	GBP (million)
14	**MINEDUC**	**149.2**	**135.1**
1421	Higher education	34.4	31.1
1422	TVET	51.0	46.2
1423	Curricula and pedagogical materials	6.3	5.7
1424	Teacher development and management	5.2	4.7
1425	Education quality and standards	10.1	9.2
1426	ICT integration	7.5	6.8
1427	Examinations and accreditation	3.5	3.2

1428	Higher education scholarship management	10.4	9.4
Districts	**Total**	**101.4**	**91.8**
of which	Pre-primary and primary education	50.0	45.2
of which	Secondary education	51.3	46.4
of which	Tertiary and non-formal education	0.2	0.2

Source: GoR MINECOFIN Budget for 2013/2014. Exchange rate used is RWF 1105 = 1 GBP, as found on 4th November 2014.

The data from the GoR budget for 2013/2014 is then used to estimate the share of the total spend going to each level of education. As shown in Table 5 this splits fairly evenly between pre-primary and primary, secondary level, TVET, and tertiary levels, with around a quarter of expenditure each. For spending categories in the MINEDUC budget not specifically classified – for example "1425: education quality and standards", we assume this to be split evenly across each different stage of education.

Table A9.5: Share of GoR education expenditure going to each level of education

Stage of education	Amount RWF (billion)	Amount GBP (millions)	% of total
Pre-primary and primary education	63.3	57.3	25.3%
Secondary education	64.7	58.5	25.8%
TVET	64.4	58.3	25.7%
Tertiary	58.3	52.8	23.3%
Total	250.7	226.9	100.0%

Source: GoR MINECOFIN Budget for 2013/2014. Exchange rate used is RWF 1105 = 1 GBP, as found on 4th November 2014. Author estimates on non-categorised spending.

In terms of the source of funds for the £227 million set out in the 2013/2014 budget, the budget indicates the amount of money coming from sector budget support (SBS) allocated to education, as well as larger allocations of general budget support (GBS). Table A9.6 shows estimates of the education spend from each source, with SBS and GBS both making up around a tenth of overall expenditure.

The final source of expenditure cited by UNESCO (2012) was off-budget project aid. Unfortunately, this is harder to find data on. Searching DFID and USAID databases we could find approximately £5.1 million of additional expenditure for the year 2013/2014. In addition there is DFID's Rwanda Education Sector Programme, which has spent an average of £22 million per year since 2011, although it is not clear if this has been largely taken into the DFID SBS allocation to the GoR budget. Given these uncertainties we have assumed that there is another £5 million or so of off-budget aid, taking the total to £10.1 million or 4 per cent of the total.

Table A9.6: Share of education expenditure coming from different sources

Source	RWF billions	GBP millions	Share of total
Education total spend	**261.8**	**236.9**	**100%**
From sector budget support (SBS) specified for education	26.7	24.2	10%
From general budget support (GBS)	20.2	18.3	8%
Tax rev, other revenue or borrowing	203.8	184.4	78%
Off-budget aid	11.2	10.1	4%

Source: GoR MINECOFIN Budget for 2013/2014. Author estimates for GBS and tax revenue shares. Off-budget aid is an estimated

Private expenditure on education

In addition to the above public funding of education there is Rwandan private expenditure on education, with the two together making up the total of education expenditure. It is difficult to find accurate estimates of private funding of education, as this constitutes not just money spent at private schools, but on education-related expenditures required even for public schooling. Additionally, evidence

suggests that parents in Rwanda also put in labour to construct classrooms, and also supplement the capitation grant to schools (Paxton and Mutesi, 2012). Parents make direct financial contributions through a parent teacher association (PTA) contribution, a sum determined by each PTA, which all parents are required to pay. The PTA contributions are used to supplement teacher salaries (Williams et al. 2014).

In 2008 it was estimated that parents met 45 per cent of the costs of education – 29 per cent at primary level; 59 per cent at junior secondary level; 68 per cent at senior secondary level; and 40 per cent at higher education level (World Bank/International Bank for Reconstruction and Development, 2011). These averages are however deceptive as parental contributions can form a substantial part of the income for schools in more affluent urban areas. The inability of parents in poor rural areas to make similar contributions has created significant inequalities in the quality of education (Paxton and Mutesi 2012).

Given the paucity of other estimates, we utilise the figures from 2008 to form our estimates of private expenditure on education.

Total expenditure on education and the share of RBA

The estimates above are assumed to represent the level of expenditure on education in Rwanda during 2012, 2013 and 2014. For reasons set out in the next section on benefits under test A, TVET has been excluded from the model due to the difficulty of accurately modelling returns to this type of education and the relatively few enrolled in TVET during 2013. As such, it is also removed from the cost expenditure. However this means that a proportion of the RBA amount equivalent to the share of GoR education expenditure going to TVET also needs to be removed.

Table A9.7: Estimated total expenditure on education in Rwanda in 2013 (excluding TVET)

	GoR expenditure		Off-budget project aid		Private expenditure		Total spend	
Stage of education	**RWF billion**	**GBP million**	**RWF billion**	**GBP million**	**RWF billion**	**GBP million**	**RWF billion**	**GBP million**
Pre-primary and primary education	63.3	57.3	-	-	25.9	23.4	89.2	80.7
Secondary education	64.7	58.5	-	-	105.5	95.5	170.1	154.0
Tertiary	58.3	52.8	-	-	38.9	35.2	97.2	87.9
Total	186.3	168.6	11.2	10.1	170.2	154.1	367.7	332.7

Source: GoR 2013/2014 budget data, author estimates for off-budget aid, World Bank (2011) data from 2008 used for estimating private expenditure.

The overall expenditure on education in 2013 in Rwanda is therefore estimated to be £332.7 million as shown in Table A9.7. The share of the total expenditure that the RBA represents (stripped of the contribution to TVET) is then 0.25 per cent in 2012, 0.40 per cent in 2013, and 0.26 per cent in 2014. These are the shares that are applied to the returns derived under test A as an estimate of the net benefit of the RBA aid. The full RBA allocation still forms the main cost comparison for the VfM final analysis for test A.

Benefits under test A

RBA has been set up with the key variable or outcome of interest as those completing, or more specifically 'sitting the final exams' at the grades of P6, S3 and S6. However, this of course forms only a small part of the outcomes within the education system more broadly, and therefore of the contribution of aid to this system.

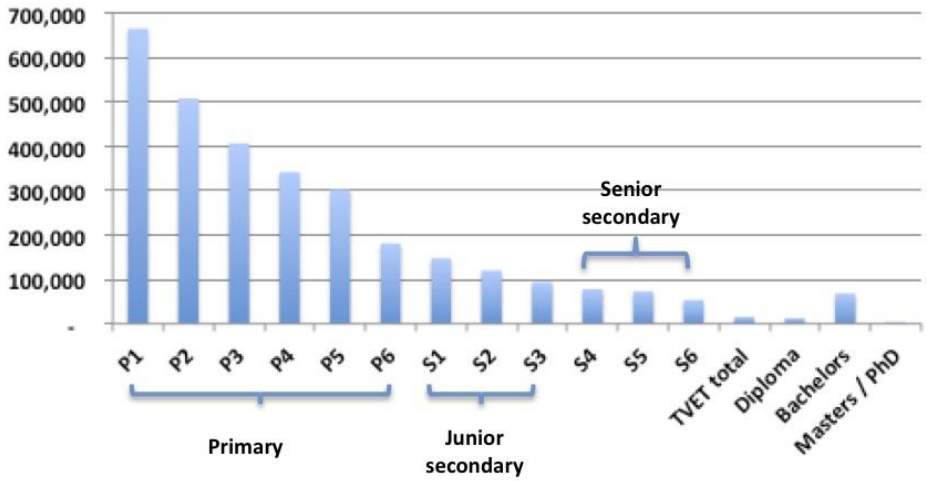
As such, outcomes or benefits need to be expressed in terms of a more clearly quantifiable variable. This comes from the standard economic tool to assess education benefits – returns to education – these are generally financial, and expressed as private returns, the additional income for an individual

per additional year of education. The challenge is then to calculate what these benefits are and express them in terms of present values.

Defining the cohort

The first challenge is to model the amount of education a given cohort of children receives. The cohort in question is the full group of children and young adults in education in 2012, as well as 2013 and 2014. For this, GoR data on the number of children in different grades and stages of education during each year provides the starting point. The cohort as defined by grade and stage of education is shown below in A9.1 for 2013, which is also almost identical to the cohort size for 2012 and 2014.

Figure A9.1: Number enrolled in each grade and stage of education in 2013



Source: GoR, 2012, 2013 and 2014 Education Statistical Yearbooks

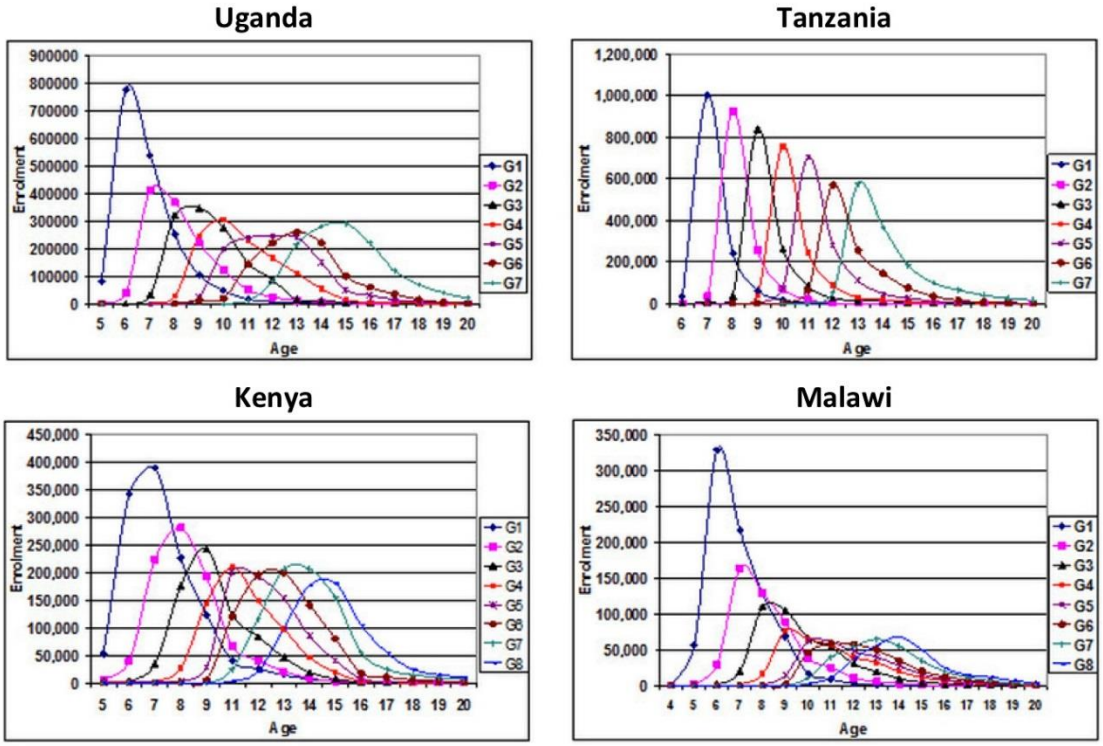
Those enrolled in technical, vocational education and training (TVET) have been excluded from the VfM model at this time. This is because the majority of TVET classes are for less than one year and therefore it is hard to count this as an additional year of education. In addition, there is no information available on how much education those enrolling in TVET courses had beforehand and at what stage they enrol, so it is hard to know whether to model it as an additional year of education for someone with three or six years of primary, or who has finished senior secondary level. For similar reasons, those taking either masters or PhD level education are also excluded from the model. The numbers enrolled in these levels are much smaller than for other levels, returns estimates are not available, and again it is not clear the age ranges for those taking these higher level qualifications.

The cohort is therefore refined to those in primary, junior secondary, senior secondary, and at tertiary level studying either for a Bachelors or Diploma qualification. These latter courses are assumed to last for four years, and are defined in the model as BD1, BD2, BD3 and BD4 for each year respectively.

Estimating the age range

The age range of a given grade is likely to be highly variable, unlike in a developed country such as the UK. As such, each grade is modelled to have a range of ages. While Rwandan data could not be found for this, regional neighbours such as Uganda and Kenya do have some data available, and as shown in Figure A9.2 there is a significant spread of ages for each grade. It is assumed that a similar pattern is likely to exist in Rwanda. Indeed the numbers of children enrolled, particularly in the early grades of primary show that this is a certainty. For example the starting age for Grade 1 of primary school in Rwanda is 7 years old, but while there were likely to be just over 300,000 children aged 7 in the country in 2013, close to 700,000 children were enrolled in Grade 1.

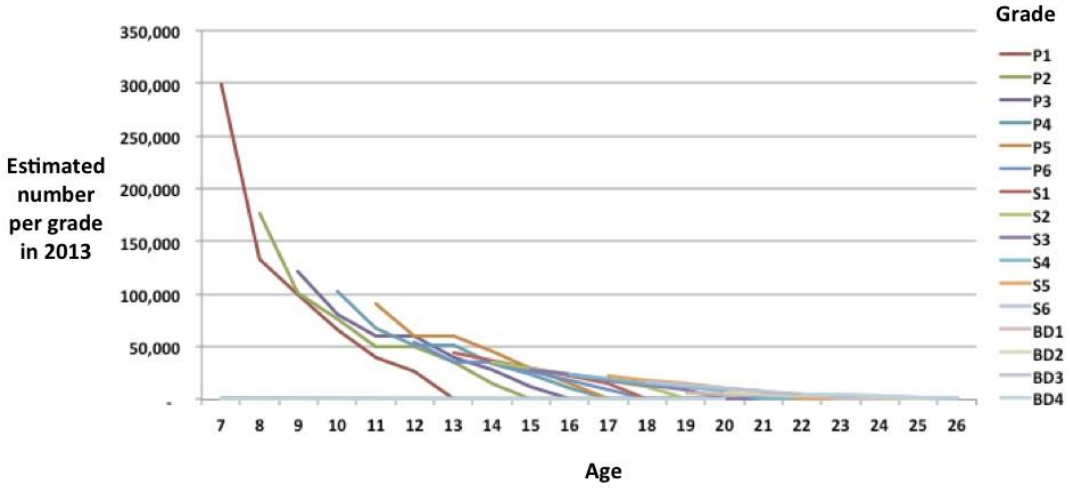
Figure A9.2: Enrolment distribution for grades by age in regional countries



Source: Lewin and Sabates (2009)

The most accurate and up-to-date population estimate in Rwanda comes from EICV3. We use this data as the estimate for 2011 population. While EICV 3 splits this by age group, we then break that down to individual ages, utilising the natural growth rate between these age groups to estimate the growth for particular ages. This information on the number of children at each age, combined with the regional estimates above, allows us to estimate the distribution of age for each grade. These are shown in Figure A9.3

Figure A9.3: Estimated breakdowns of each grade by age utilised for VfM model



Source: Author estimates, based on information from NISR (2012) and Lewin and Sabates (2009)

Estimating the current numbers per year group for tertiary level

The tertiary level information available does not break down the overall number of students into the individual year groups. Given the bachelors courses in Rwanda take four years, we estimate the

numbers split between four years taking bachelors or diploma courses. This is done utilising an assumed drop out per year of 6 per cent; assumed, as this is the drop-out rate available for upper secondary level.

Table A9.8: Split between different years of tertiary level assumed in VfM model

Stage of education	2012	2013	2014
Number enrolled for Bachelors qualification	63,111	68,246	71,084
Number enrolled for Diploma qualification	10,984	12,557	11,542
Bachelors and Diploma number combined	**74,095**	**80,803**	**82,626**
Number assumed in first year: BD1	20,173	21,999	22,495
Number assumed in second year: BD2	19,031	20,754	21,222
Number assumed in third year: BD3	17,954	19,579	20,021
Number assumed in fourth year: BD4	16,938	18,471	18,887

Source: GoR (2014b). Author estimates. Assumption of 6 per cent annual drop-out.

Drop out, repetition and progression

The VfM model in 2014 utilised the data provided by the GoR 2013 Education Statistical Yearbook for estimates of the drop-out rates, repetition rates and progression rates for particular grades and between different levels of education (GoR, 2014b). For the 2015 model, data from the years 2012, 2013 and 2014 are used for estimating drop-out rates, repetition rates and progression rates. The data is used to estimate a range of estimates with a 'low drop-out and repetition rate' scenario using the lowest figures from these years, and a 'high drop-out and repetition rate' scenario using the highest figures. The average of these two figures is then used in the final model.

Unfortunately no data is available on drop-out rates for tertiary level education, so these are assumed to be the same as for secondary. The rate of transition from finishing upper secondary to tertiary education is estimated using the estimate from the cohort and is kept at 41.1 per cent for both scenarios. The final rates used are shown in Table A9.9. All of these rates are then assumed to carry on into the future.

Table A9.9: Repetition, drop out and progression rates used in VfM model

	Repetition rate	Drop-out rate	Progression rate
Primary level (low estimates)	12.5%	10.9%	76.4%
Lower secondary level (low estimates)	5.8%	13.1%	81.2%
Upper secondary level (low estimates)	1.7%	2.5%	95.9%
Upper secondary to tertiary transition	-	-	41.1%
Tertiary level (low estimates)	-	2.5%	97.6%
Primary level (high estimates)	18.2%	14.2%	67.6%
Lower secondary level (high estimates)	11.7%	17.7%	70.7%
Upper secondary level (high estimates)	3.8%	6.2%	90.1%
Tertiary level (high estimates)	-	17.7%	82.3%

Source: GoR (2014b), progression rate from upper secondary to tertiary calculated using 2012 and 2013 data, tertiary level drop-out rate assumed to be same as for upper secondary. Percentages that do not add up to 100 per cent for a given row is due to rounding error.

Running the education model into the future

The model is used to run through into the future to see where and when different cohorts leave education, which in turn is used to estimate the years of education gained. Years of education are calculated based on the highest grade an individual reaches – so for example a child dropping out

during grade 5, will be assumed to have had four full years of education, regardless of if they repeated a particular grade, and are then classified accordingly. In total there are just over 3 million individuals in the cohort for each of the three years, and the model is run for these individuals until 2033 for the 2012 cohort, 2034 for the 2013 cohort, and 2035 for the 2014 cohort, based on all of the key assumptions laid out above, notably that repetition, drop-out and progression rates continue and do not decline or increase over time. Twenty years is a period that allows the initial P1 cohort to get to graduation at BD4 with some repetition.

Modelling the labour market

The next step is to then model the returns to education accrued by the cohort during their working lives. The model extends for the full working lives of the current cohort, assuming that for those not still in education, paid employment begins at the age of 16. The model then runs up to retirement, which is assumed to be the current Rwanda retirement age of 60 years old[79], taking the youngest part of the cohort, aged 7 in 2014, through to their retirement in 2067.

As discussed above, the 2015 model makes some significant additions to ensure greater realism of the model. This includes adding mortality rates, taking into account periods of unemployment and economic activity, assuming them to be 4.4 per cent as they were in EICV3, and real wage growth rates of 4 per cent, again based on EICV3 and the growth in consumption experienced from 2006 to 2011 (the annual rate of change between the EICV2 and EICV3 surveys).

Returns to education

The most important variable in generating benefits in the model is the economic return to schooling assumed to accrue for every additional year of education. As discussed above, these are changed for the 2015 model based on a broader and deeper assessment of the data available. This is set out in section 2.1 above so is not repeated here. The returns to education estimates are applied to the wage rate for someone with no education, which is estimated based on the WageIndicator data as well as the triangulation of this EICV discussed above to be RWF153,000 for 2013.

Table 10: Returns estimates used in the 2015 model and comparison to 2014 model

Level of education	Return to year of schooling used in 2014 model	Return to year of schooling used in 2015 model
Primary	15%	10%
Lower Secondary	15%	15%
Upper Secondary	20%	17.5%
Tertiary	20%	25%

Rates of return to schooling and the starting wage are used in the model to calculate the wages in the labour market for each individual in the cohort for any given year. If they are in employment in that year – which means the have left education, are aged 16 years or older, aged 60 years or younger, are not unemployed or economically inactive, or succumbed to mortality – then their wages are added to the net benefits of the model.

The share of the returns to education for which one year is responsible

One of the clear complicating factors of estimating the life-time returns to a cohort for a given year of education is that the process of education takes many years. As such this year's investment in education is just a small component of the total amount received for any given child. The model therefore needs to calculate how much of the returns measured are attributable to this year's education. In order to calculate this, a measure is required that shows the number of years of education provided to all those in the cohort.

[79] http://www.csr.gov.rw/content/pension-benefits

The model estimates this into the future, including repetition. Into the past, the model assumes that each individual completed all the grades prior to their current grade, without repetition. Completing this modelling shows for each grade the percentage of the overall education received that occurred this year. This percentage is then applied to the overall wages of the cohort, and is seen as the attribution of the overall benefits to the year in question.

Discount rate applied

As the model seeks to measure returns far into the future, a discount rate is essential for the model to accurately reflect the measure of costs and benefits faced today. As set out in White, Hodges and Greenslade (2013) and the HM Treasury Green book (HM Treasury, 2011), there are two approaches to thinking about the discount rate:

- **the social rate of time preference (SRTP),** which assesses the value society attaches to present as opposed to future consumption; and
- **the social opportunity cost of capital (SOC)** which seeks to proxy the marginal social return if funds were invested privately; this is seen as more appropriate in certain developing country contexts with severe resource scarcities – including limited access to international finance markets – which mean that SRTP will understate the 'true' discount rate.

The SRTP is likely to be higher in developing countries, due to the expectation that higher growth rates will occur in future combined with a declining marginal utility of income. This means the future should be discounted at a higher rate than in a lower growth country. The discount rate applied in the model is therefore 10 per cent, which also follows DFID guidance of a discount rate in the range of 8-12 per cent for developing countries (p.24 of "Guide to Investment Appraisal for DFID Economists", 2005). The sensitivity analysis undertaken below also looks at changes if the discount rate differs.

Overall results for Test A

The overall results for Test A find the PV of benefits to be £11.2 million for the total of 2012, 2013 and 2014. This is significantly lower than the estimate in the 2014 report, which is due to the multiple changes in assumptions detailed above. The PV of costs is the combined disbursements of £1.2 million, £1.9 million and £1.2 million disbursed in RBA for the 2012, 2013 and 2014 results, with the first two years adjusted for UK inflation to ensure that the results are all reported in 2015 GBP. This implies a NPV of £6.8 million, and a benefit-cost (B-C) ratio of 2.6. This suggests that aid to education does provide good overall VfM, even when significantly discounting future returns to education. The NPV and B-C ratio are lower than in the 2014 model, which again is due to the more stringent assumptions applied for the 2015 modelling.

Sensitivity analysis

In order to look at how sensitive the model developed is to variations in the assumptions made, Table A9.11 shows the NPVs and benefit-cost ratios for twelve different scenarios:

1. The discount rate at the lower end of the DFID recommended range at 8 per cent, rather than 10 per cent.
2. The discount rate at the upper end of the DFID recommended range at 12 per cent.
3. Higher rate of transition from upper secondary to tertiary level education – 20 per cent higher at 61.1 per cent.
4. Lower rate of transition from upper secondary to tertiary level education – 20 per cent lower at 21.1 per cent.
5. A scenario with higher returns to education, based on some of the higher estimates discussed above in this paper – with 12.5 per cent returns for primary, 17.5 per cent for lower secondary, 20 per cent for upper secondary, and 27.5 per cent returns to education for tertiary level.
6. A scenario with lower returns to education – with 7.5 per cent returns for primary, 12.5 per cent returns for lower secondary, 15 per cent for upper secondary, and 22.5 per cent returns for tertiary level.

7. Higher wage growth –the rate of real wage growth in Rwanda at 6 per cent rather than 4 per cent.
8. Lower wage growth - the rate of real wage growth in Rwanda at 2 per cent.
9. Lower levels of unemployment and economic inactivity – at 0 per cent.
10. Higher levels of unemployment and economic inactivity – at 20 per cent.
11. Money spent on aid to education worth more than the average pound (due for example to extra influence from donors on improving policy), at £1.25 for every other £1.
12. Money spent on aid to education worth less than the average pound (due for example to fungibility), at £0.75 for every other £1.

Table 11: Sensitivity analysis for test A (GBP million, except for B-C ratio)

Scenario	PV of benefits	Cost of RBA	NPV	Benefit-cost ratio
Overall results for Model A	11.2	4.4	6.8	2.6
1) Discount rate lower (8%)	15.8	4.4	11.5	3.6
2) Discount rate higher (12%)	8.2	4.4	3.9	1.9
3) Transition from upper secondary to tertiary increases (to 61.1%)	11.8	4.4	7.5	2.7
4) Transition from upper secondary to tertiary decreases (to 21.1%)	10.5	4.4	6.1	2.4
5) Greater rates of return to schooling (12.5% primary, 17.5% lower secondary, 20% upper secondary, 27.5% tertiary)	15.7	4.4	11.3	3.6
6) Lower rates of return to schooling (7.5% primary, 12.5% lower secondary, 15% upper secondary, 22.5% tertiary)	7.7	4.4	3.3	1.8
7) Higher real wage growth (6%)	16.3	4.4	12.0	3.8
8) Lower real wage growth (2%)	7.9	4.4	3.5	1.8
9) Lower levels of unemployment and economic inactivity (0%)	11.7	4.4	7.3	2.7
10) Higher levels of unemployment and economic inactivity (20%)	9.3	4.4	5.0	2.1
11) Money spent on education through aid worth more than other expenditure (X1.25)	14.0	4.4	9.6	3.2
12) Money spent on education through aid worth less than other expenditure (X0.75)	8.4	4.4	4.0	1.9

The analysis here shows that the model is of course most sensitive to changes in the rates of returns to education, which generate the benefits for this model. The costs assessed remain constant as they have been assumed to be the RBA disbursed. There may be reasons why real wage growth would not stay at 4 per cent as used in the model over time, particularly if Rwandan economic growth were to slow, and/or be predominantly based in low employment industries such as those oriented to export or service industries based in Kigali, which would not necessarily benefit the average Rwandan. As such the sensitivity case for lower real wage growth should be considered a possibility.

In addition to this, it should be noted that returns to education figures do generate some controversy in the economics literature. In particular, the distinction between quantity and quality is often seen as a critical distinction. While there is little data available for developing countries due to the paucity of data on quality measures, this finding should be taken into account, given that the model presented here is very much focussed on quantity rather than quality.

One note of caution on this result should perhaps be on the assumption that all expenditure on education is equal. For various reasons, it may be that money spent by private individuals is more likely to generate returns to education, this could be because individuals are better able to make educational choices, and because individuals are less likely to waste their own resources. If private expenditure on

education generated a greater proportion of the returns, then the returns estimated in model A would be lower as shown in the sensitivity analysis – if it a pound spent by aid is worth 75 per cent of a pound spent by Government or privately, the B-C ratio reduces from 2.6 to 1.9. This should be a pointer towards further research around the fungibility of aid, referring to the likes of public expenditure tracking surveys.

TEST B: VfM against counterfactual of aid provided by other means

The second test for VfM seeks to look at the additional value of providing RBA against the counterfactual of still providing the aid but via another means. The other aid means is assumed to be a non-PbR aid modality without any specific incentive effect, for example by sector budget support (SBS). This is the VfM test of the more direct theory of change of RBA – that an incentive effect leads to greater performance against the defined metrics – in this case the completion rates for P6, S3, and S6 – and that whether this outweighs any additional costs of the RBA instrument.

The VfM approach used distinguishes between two ways of looking at test B, the first B1 looks at the model at face-value, that is utilising the payment mechanism design where the underlying time-trend for change is effectively zero, and any positive change is counted. This means looking at the number of 'extra sitters' of examinations at P6, S3 and S6, assuming these to be caused by RBA, and then measuring the results into the future.

The second version of test B, B2 looks at the RBA evaluation's econometric model, and the specific predictions of that model for how much of the change measured can be said to be additional. Thus, 'additional sitters' are used for B2, and modelled to measure benefits into the future. A major change for the 2015 report and model is that for B2 statistically significant negative changes are assessed as well as positive changes, and as expected this significantly reduces the net benefits of the model. The 2015 report also models for 2012, 2013 and 2014 changes.

Costs under test B

The costs under test B differ from those for test A. The counterfactual for test B is that the *aid would have been provided but by another means – therefore the aid itself is not a cost for test B*. The costs for test B are instead the specific costs associated with RBA that would not be present under another instrument. This includes the cost of verification and the cost of the evaluation of the RBA pilot.

The costs therefore include the £90,000 for the cost per year for verifying the 2012, 2013 and 2014 results, totalling £270,000. In addition, the evaluation cost of the RBA pilot, which comes to £462,293. Both the verification and evaluation costs are assumed to be additional costs of using the RBA instrument, and are therefore counted as a cost for both test B1 and B2.

In addition to these costs, there are costs associated with the model for test B. These costs are not accrued by DFID but by the education system itself, and therefore by the Government of Rwanda as well as household expenditure on education. The model estimates the numbers of years additional completers at each grade will continue in education. The cost of this education is then attributed to the fact they have now progressed through school (as in the counterfactual they would have left school).

The costs used by the model are those using GoR (2014a) budget data, as well as GoR (2014b) education data. This finds unit costs per person at different levels of education – these costs are £32 per primary school pupil per year, £272 per secondary school pupil per year, and £1,041 per tertiary pupil per year. These costs are uprated by an assumed inflation rate of 5 per cent per year to account for the costs demonstrated by the model. Table A9.12 shows the high level PV of these costs for different types of completers.

Table 12: Present value of costs for additional completers at P6, S3, S6 (GBP millions)

Level of completion	Present value of associated costs (£ millions)
1,000 additional students complete P6	1.86
1,000 additional students complete S3	2.01
1,000 additional students complete S6	1.58

Benefits under test B

The benefits under test B will be those ascribable to the presence of RBA as an incentive mechanism, as measured by the payment mechanism itself (for B1) and by the Upper Quartile econometric component of the RBA evaluation (for B2). The key assumption here is that what happens following the introduction of RBA would not have happened anyway. There are two ways of modelling this assumption, providing two estimates for Test B, these are:

B1: The benefits from the 'extra' completers at P6, S3 and S6, as paid out upon through the payment mechanism of RBA as it has been designed.

B2: The benefits from 'additional' completers at P6, S3 and S6, estimated to be statistically significant by the econometric modelling of the evaluation. For the 2015 model the costs from statistically significant reductions in completers for given years are also included.

The B1 estimate of benefits here is clearly less rigorous. This is because the baseline time-trend for the RBA payment mechanism is assumed to be flat, so any increase experienced is paid out upon, while negative changes are ignored. The counterfactual of a flat time-trend is not particularly scientific given population growth alone, although it is the counterfactual on which the mechanism has been designed. This is generous to RBA in the sense that it is impossible to reject the hypothesis that improvements happened without RBA, it does provide a benchmark with few assumptions though.

The second benefit calculation is much more rigorous and is based on looking at the underlying time-trend of previous years, and other more rigorous assumptions. Completers of P6, S3 and S6 will only be considered additional if they can be shown to be by the econometric modelling used in the evaluation. For the 2015 model, negative statistically significant changes are also modelled. The numbers of additional sitters for each test are shown in Table A9.13, which shows that the statistically significant changes for B2 have been as often negative as positive.

Table 13: Number of extra and additional sitters used for model B1 and model B2

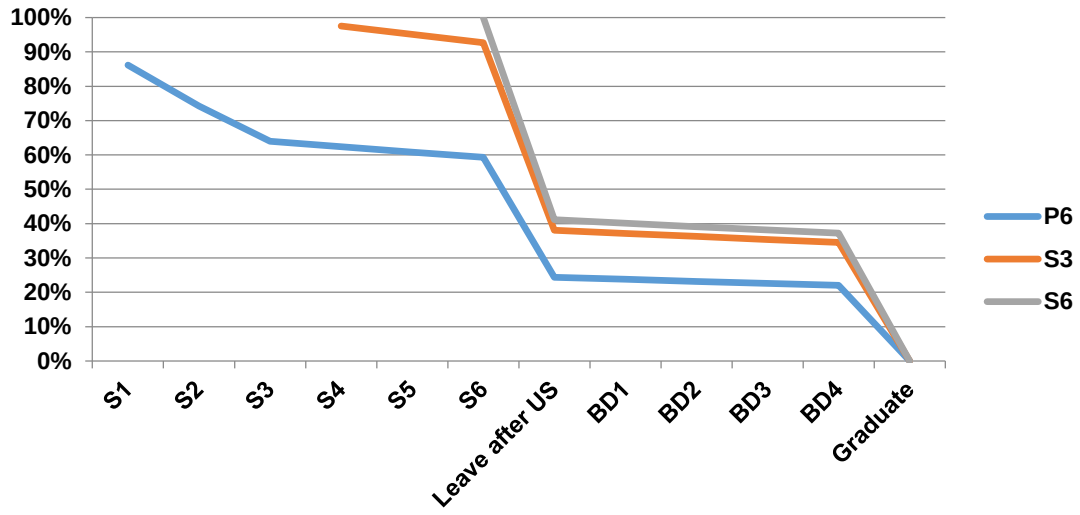
	B1: Extra sitters (as paid out upon by RBA)			B2: Statistically significant change in sitters (modelled by econometric report)		
Year	**2012**	**2013**	**2014**	**2012**	**2013**	**2014**
Additional students complete P6	11,199	0	0	0	-3,636	-14,313
Additional students complete S3	3,170	13,142	0	-2,127	10,233	0
Additional students complete S6	3,744	3,489	15,492	-8,208	3,273	13,584

Estimating additional years of education achieved

The three grades of interest for the RBA model as designed are P6, S3 and S6. The key step to estimate the benefits of this model of RBA is therefore to estimate the additional years of education that are likely to occur for the extra or additional completers for these grades. The methodology applied is exactly the same for both the B1 and B2 methodologies.

The same model used for test A was applied here. This includes assumptions about the repetition, drop-out and progression rates, taken from GoR data. Figure A9.4 shows the path through education for additional or extra completers at each of the three grades of interest given the model's estimates of drop-out and repetition rates. This shows a sharp decline for the large numbers of students who do not progress to tertiary level education despite completing secondary school (only 41.1 per cent of those finishing S6 are assumed to go on to tertiary level).

Figure A9.4: Proportion of cohort still in education at each level for extra sitters at P6, S3 and S6[80]



Modelling labour market outcomes for test B

The assumptions underlying test A are again applied to test the labour market outcomes for test B, where each additional completer at P6, S3 and S6, will have a greater number of years of education leading to higher wages when they are in employment. Notably this includes the assumption that individuals no longer in school enter employment at 16 years of age, and that all individuals end work at 60 years of age. In addition the same returns to education estimates are used as for test A. As with test A for the 2015 model additional factors have been added to make the model more realistic including mortality rates and rates of unemployment and economic inactivity.

The model must then measure the benefits of additional progression at P6, S3 and S6 in the labour market as returns to education. This is measured against the counterfactual of them dropping out of education at these stages. Crucially, these individuals would also be earning income in the labour market, but less due to the lower returns to education with less years of schooling. Table A9.14 shows the PV of the benefits for different levels against the counterfactual. These show that the PV of earnings are higher for those progressing at higher levels of education, which reflects the greater number of years of schooling they achieve. However, they also would earn more anyway as represented by the counterfactual of their dropping out – both because they still have more years of schooling at S6 than S3, and at S3 than P6, but they also enter the labour market sooner than their younger peers – and so the value of earnings is not discounted to the same degree. Overall the PV of the net returns is highest for those additionally progressing at S3.

[80] Figure shows how much of the cohort progressing for each grade remains at each subsequent year of education. S1-S3 are lower secondary grades 1 to 3, and S4-S6 are upper secondary grades 4 to 6. US stands for upper secondary. BD1-BD4 are the year of undergraduate education respectively.

Table A9.14: Present value of returns to education for additional progression at P6, S3 and S6

Cohort	Counterfactual	PV of returns to education (lifetime earnings) if continuing in school (£ million)	PV of lifetime earnings in counterfactual of leaving school (£ million)	PV of cohort less counterfactual (Change in net lifetime earnings) (£ million)
1,000 additional progressing at P6	1,000 dropping out at P6	6.4	3.5	3.0
1,000 additional progressing at S3	1,000 dropping out at S3	9.9	5.4	4.5
1,000 additional progressing at S6	1,000 dropping out at S6	12.6	8.5	4.1

Figures that do not add up for a given row due to rounding error.

Extrapolating using 2012, 2013 and 2014 results

The above methodology is then applied to the actual numbers from 2012, 2013 and 2014 shown above in Table 15. In terms of the test for B1, this is simply the extra number of sitters at P6, S3 and S6 (with negative numbers counted as zero). For B2, this is the 'additional' number, i.e. the statistically significant measure of change. This draws on the evaluation econometric report and model. Specifically the model 2b is used from that report. As noted above, for this 2015 report, statistically significant negative changes are used as well as statistically significant positive changes.

Overall findings of test B1

The test B1 finds very large benefits, with the 13,000 extra sitters at S3 and over 3,000 sitters at S6 generating a PV of benefits of over £160 million. This is offset against the additional costs of this education of over £30 million, as well as the small evaluation and verification costs, generating a NPV for test B1 of over £130 million, with a benefit-cost ratio of 5.3. While built on a large number of assumptions, these returns would suggest that the RBA appears to be very good VfM. However, this is only if these sitters can be viewed as additional to a counterfactual of providing aid through another mechanism. For test B1 this finding is not likely to be robust, so the large finding here should probably be discounted relative to test B2.

Table A9.15: Overall findings summary for test B1 (GBP millions)

Year of change	PV benefits	PV costs	NPV	B/C ratio
2012	65.6	34.7	30.8	1.9
2013	74.5	33.4	41.2	2.2
2014	63.2	24.5	38.7	2.6
Evaluation and verification costs		0.7		
Total	203.3	93.4	109.9	2.2

Overall findings of test B2

The overall findings for the test B2 are much lower than for test B1, but is much more robust, as the numbers are built on econometric modelling designed to calculate how much of the change can be viewed as additional. However the impact of the negative changes being added to the model very much weighs on the model, with the negative changes in 2012 in particularly having to be off-set by the statistically significant positive changes to completion in 2013 and 2014.

Table A9.16: Overall findings summary for test B1 (GBP millions)

Year of change	PV benefits	PV costs	NPV	B/C ratio
2012	21.1	52.1	-31.0	0.4
2013	72.7	41.0	31.7	1.8
2014	82.0	64.1	17.9	1.3
Evaluation and verification costs		0.7		
Total	175.8	157.9	17.9	1.1

A note of caution to the findings for both B1 and B2 is that the model assumes stability in underlying variables. This includes the repetition, drop-out and progression rates discussed above, but also includes the quality of education. If it was the case that those students who were now additionally sitting the final tests for P6, S3 and S6, were of a lower quality than other students, then they may face higher drop-out rates in future and therefore be less likely to progress through education as the model would predict.

In addition, the VfM findings here should very much be taken together with the econometric findings, and the qualitative findings from the overall evaluation of RBA. The findings for test B in particular are only really relevant if the overall evidence suggests that the GoR are responding to the incentive mechanism provided by RBA, and in a way that does not mean that perverse incentives have been generated and the general quality of education has diminished.

Sensitivity analysis

Given the scale of the model used to assess the VfM of RBA, it is important not to ignore the large part played by some of the key assumptions around model variables. We have therefore conducted sensitivity analysis to observe how the value created by the model responds to changes in the variables. Notably for test B, we have looked at ten additional scenarios:

1. The discount rate at the lower end of the DFID recommended range at 8 per cent, rather than 10 per cent.
2. The discount rate at the upper end of the DFID recommended range at 12 per cent.
3. Higher rate of transition from upper secondary to tertiary level education – 20 per cent higher at 61.1 per cent.
4. Lower rate of transition from upper secondary to tertiary level education – 20 per cent lower at 21.1 per cent.
5. A scenario with higher returns to education, based on some of the higher estimates discussed above in this paper – with 12.5 per cent returns for primary, 17.5 per cent for lower secondary, 20 per cent for upper secondary, and 27.5 per cent returns to education for tertiary level.
6. A scenario with lower returns to education – with 7.5 per cent returns for primary, 12.5 per cent returns for lower secondary, 15 per cent for upper secondary, and 22.5 per cent returns for tertiary level.
7. Higher wage growth – with the rate of real wage growth in Rwanda at 6 per cent rather than 4 per cent.
8. Lower wage growth - the rate of real wage growth in Rwanda at 2 per cent.
9. Lower levels of unemployment and economic inactivity – at 0 per cent.
10. Higher levels of unemployment and economic inactivity – at 20 per cent.
11. Lower rate of cost inflation (applied to schooling costs) – at 2.5 per cent.
12. Higher rate of cost inflation (applied to schooling costs) – at 10 per cent.

The results of the scenario analysis for the different benefits of P6, S3 and S6 completers is shown below in Table A9.17. This shows that the variables of the discount rate and the rate of return to education have very large impacts on the measured benefits from the model. Changes to the average

rate of drop-out, as well as transition from upper secondary to tertiary level education also see significant changes to the PV of benefits estimated by the model.

Table A9.17: Sensitivity analysis on results of the model test B1, (GBP millions)

Scenario	PV benefits	PV costs	NPV	B/C ratio
Overall	203.3	93.4	109.9	2.2
1) Discount rate lower (8%)	292.7	97.9	194.8	3.0
2) Discount rate higher (12%)	144.7	89.3	55.4	1.6
3) Transition from upper secondary to tertiary increases (to 61.1%)	253.1	124.3	128.7	2.0
4) Transition from upper secondary to tertiary decreases (to 21.1%)	153.6	62.5	91.2	2.5
5) Greater rates of return to schooling (12.5% primary, 17.5% lower secondary, 20% upper secondary, 27.5% tertiary)	307.2	93.4	213.8	3.3
6) Lower rates of return to schooling (7.5% primary, 12.5% lower secondary, 15% upper secondary, 22.5% tertiary)	129.8	93.4	36.4	1.4
7) Higher real wage growth (6%)	302.6	93.4	209.2	3.2
8) Lower real wage growth (2%)	138.3	93.4	44.9	1.5
9) Lower levels of unemployment and economic inactivity (0%)	212.7	93.4	119.3	2.3
10) Higher levels of unemployment and economic inactivity (20%)	170.2	93.4	76.8	1.8
11) Lower rate of cost inflation (applied to schooling costs) – at 2.5 per cent.	203.3	85.9	117.4	2.4
12) Higher rate of cost inflation (applied to schooling costs) – at 10 per cent.	203.3	110.7	92.7	1.8

Table A9.18: Sensitivity analysis on results of the model test B2, (GBP millions)

Scenario	PV benefits	PV costs	NPV	B/C ratio
Overall	175.8	158.0	17.9	1.1
1) Discount rate lower (8%)	229.4	209.9	19.6	1.1
2) Discount rate higher (12%)	139.5	124.0	15.5	1.1
3) Transition from upper secondary to tertiary increases (to 61.1%)	222.6	201.0	21.6	1.1
4) Transition from upper secondary to tertiary decreases (to 21.1%)	129.1	114.9	14.2	1.1
5) Greater rates of return to schooling (12.5% primary, 17.5% lower secondary, 20% upper secondary, 27.5% tertiary)	237.6	211.9	25.8	1.1
6) Lower rates of return to schooling (7.5% primary, 12.5% lower secondary, 15% upper secondary, 22.5% tertiary)	132.3	119.3	13.1	1.1
7) Higher real wage growth (6%)	231.8	213.3	18.5	1.1
8) Lower real wage growth (2%)	138.6	122.4	16.2	1.1
9) Lower levels of unemployment and economic inactivity (0%)	181.4	162.9	18.5	1.1

10) Higher levels of unemployment and economic inactivity (20%)	156.1	140.6	15.6	1.1
11) Lower rate of cost inflation (applied to schooling costs) – at 2.5 per cent.	170.7	154.4	16.4	1.1
12) Higher rate of cost inflation (applied to schooling costs) – at 10 per cent.	188.0	166.0	22.0	1.1

The key thing to note about the sensitivity analysis is that it shows that the model is sensitive to variations in a number of variables. The model has needed to make a number of assumptions, often this constitutes an assumption that current trends will continue into the future. In reality change can be discontinuous. Particularly with respect to RBA, it is very important to note again that the assumption that RBA itself is not a discontinuity is key – i.e. RBA does not change behaviour around other aspects of the system that would impact on educational outcomes. This is particularly relevant with respect to the quality of education, which is likely to be the main underlying driver of economic and non-economic returns to education. The models presented here can only be meaningful if quality is either stable or improving over time.

Cost effectiveness per sitter - VfM findings

Given all of the assumptions noted, it is instructive to look at the VfM for the changes that RBA intends to bring about. Table A9.19 provides the unit breakdowns for test B2 in terms of how much benefit is derived from an *additional sitter* at each of the grades, P6, S3 and S6. This is informative more generally in terms of the scale of change that is required to make RBA good VfM. It shows that the NPV per individual additional sitter ranges from around £1,000 to £2,500, with this benefit the additional lifetime earnings as a result of extra years of schooling, offset by the cost of that schooling.

The result of this is that a relatively low number of additional sitters are required to break-even on the additional costs from the RBA mechanism – the evaluation and verification costs. Thus, 493 additional individuals completing P6 would be the break-even point, or 228 at S3, or 221 at S6. This shows that if RBA can be shown to have a significant and positive incentive effect, then even at a relatively small scale it should be VfM from the point of view of the cost-effectiveness model generated for this paper.

Table A9.19: Analysis per additional sitter for test B2

Additional sitter at Grade:	PV benefits	PV costs	NPV per individual additional sitter	Number of sitters required to break-even on additional RBA costs (evaluation + verification)
P6	£2,976	£1,856	£1,120	493
S3	£4,499	£2,079	£2,419	228
S6	£4,080	£1,584	£2,496	221

Bibliography

Becker, Greg (2013) What rate of wage growth can I expect? - http://www.slideshare.net/greg_becker/gg-slide-share-what-rate-of-wage-growth-can-i-expect

Besamusca, J., Tijdens, K.G., Ngeh Tingum, E., Mbassana E.M. (2013) Wages in Rwanda. Wage Indicator survey 2012. Amsterdam, Wage Indicator Foundation, Wage Indicator Data Report.

Education Policy and Data Centre 2014 Rwanda Update - http://www.epdc.org/sites/default/files/documents/EPDC%20NEP_Rwanda.pdf

Fox, L. and A.M. Oviedo (2008). 'Are Skills Rewarded in Sub-Saharan Africa'. World Bank Policy Research Working Paper 4688. Washington, DC: World Bank.

Gakidou, E., K. Cowling, R. Lozano, and C. Murray. (2010). "Increased Educational Attainment and its Effect on Child Mortality in 175 Countries between 1970 and 2009: A Systematic Analysis." The Lancet 376 (9745): 959–74.

Garcia M, Fares J. (2008) 'The effect of education on income and employment.' In: *Youth in Africa's Labor Market*. Garcia M, Fares J, editor. Washington, DC: The World Bank; 2008. pp. 39–47.

Golub, Stephen and Faraz Hayat (2014) Employment, unemployment, and underemployment in Africa. WIDER Working Paper 2014/014

GoR (2014a) MINECOFIN Budget for 2013/2014

GoR (2014b) 2013 Education Statistical Yearbook

Heckman, James J. & Kautz, Tim, 2012. "Hard evidence on soft skills," Labour Economics, Elsevier, vol. 19(4), pages 451-464.

HM Treasury (2011) Green Book 2003, 2011 update - https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/220541/green_book_complete.pdf

Human Resource Development Agency (HIDA) (2009) 'National Skills Audit Report.' Ministry of Public Services and Labour. Government of Rwanda.

International Bank for Reconstruction and Development/World Bank (2011). Rwanda Education Country Status Report: Toward Quality Enhancement and Achievement of Universal Nine Year Basic Education; an Education System in Transition; A Nation in Transition. Washington: International Bank for Reconstruction and Development/World Bank.

Kingdon, Geeta and Mans Söderbom (2008a) 'Education, skills, and labor market outcomes: evidence from Ghana.' Education working paper series, No.12. Washington, DC: World Bank.

Kingdon, Geeta and Mans Söderbom (2008b) 'Education, Skills, and Labor Market Outcomes: Evidence from Pakistan.' Education Working Paper Series. Vol. 11. Washington, DC: The World Bank.

Lassibille, G. and Tan J-P (2005) The Returns to Education in Rwanda, Journal of African Economies, 14(1), pp. 92-116.

Lewin, Keith M. and Ricardo Sabates (2009) 'Who Gets What? Is Improved Access to Basic Education Pro-Poor in SSA?' CREATE-RPC.

MIFOTRA (2013) EICV3 Thematic Report: Socio-economic status of Youth

MINECOFIN and NISR (2014) Fourth Population and Housing Census 2012. Thematic Report: Mortality. Ministry of Finance and Economic Planning National Institute of Statistics of Rwanda

National Institute of Statistics of Rwanda (2012) Integrated Household Living Conditions Survey 3 (EICV 3)

New Times (2015) 'Report: Nearly half of masters degree holders underemployed'. Solomon Asaba. January 27, 2015. - http://www.newtimes.co.rw/section/article/2015-01-27/185350/

Paxton, W., and L. Mutesi (2012) School Funding and Equity in Rwanda. Kigali:Institute of Policy Analysis and Research.

Private Sector Federation (PSF) and OTF Group (2008) Business Investment Climate Survey

Psacharopoulos, G. and Patrinos, H.A. (2004) 'Returns to Investment in Education: A Further Update'. Education Economics 12(2): 111-134.

Schultz, T.P. (2004) 'Evidence of Returns to Schooling in Africa from Household Surveys: Monitoring and Restructuring the Market for Education'. Journal of African Economies 13(2): 95-148.

Söderbom, M., F. Teal, and A. Wambugu (2002). 'Does Firm Size Really Affect Earnings?'. CSAE WPS/2002-08. Oxford: University of Oxford.

UNESCO (2012) 'Untangling Aid in National Education Budgets'. Technical Note prepared for the Education for All Global Monitoring Report 2012.

White, Philip, Anthony Hodges and Matthew Greenslade (2013) Guidance on measuring and maximising value for money in social transfer programmes – second edition. DFID.

Williams, Timothy P., Pamela Abbott, and Alfred Mupenzi (2014) 'Education at our school is not free': the hidden costs of fee-free schooling in Rwanda.' Compare. Routledge.

VfM Assessment Annex 1: Additional tables

Table A9.20: Cohort split assumed for each grade by age

Grade	Cohort size	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26
P1	664161	45%	20%	15%	10%	6%	4%														
P2	507092		35%	20%	15%	10%	10%	7%	3%												
P3	406487			30%	20%	15%	15%	10%	7%	3%											
P4	341454				30%	20%	15%	15%	10%	7%	3%										
P5	301957					30%	20%	20%	15%	10%	5%										
P6	181013						30%	20%	20%	15%	10%	5%									
S1	147547							30%	25%	20%	15%	10%									
S2	120001								30%	25%	20%	15%	10%								
S3	93974									30%	25%	20%	15%	10%							
S4	78300										30%	25%	20%	15%	10%						
S5	73026											30%	25%	20%	15%	10%					
S6	53522												30%	25%	20%	15%	10%				
BD1	21999													30%	25%	20%	15%	10%			
BD2	20754														30%	25%	20%	15%	10%		
BD3	19579															30%	25%	20%	15%	10%	
BD4	18471																30%	25%	20%	15%	10%

Table A9.21: Calculating the share of years for Model 1 for which education in 2013 is responsible

Grade	Person-years of education pre-2013	Person-years of education 2013 or later	Person-years of education Total	Person-years of education in 2013	2013 % of total
P1	2,385,177	759,041	3,144,218	664,161	21.1%
P2	1,878,085	1,242,285	3,120,370	507,092	16.3%
P3	1,471,598	1,549,249	3,020,846	406,487	13.5%
P4	1,130,144	1,742,949	2,873,092	341,454	11.9%
P5	828,187	1,866,937	2,695,124	301,957	11.2%
P6	647,174	1,836,975	2,484,148	181,013	7.3%
Primary average			**17,337,798**	**2,402,164**	**13.9%**
S1	499,627	1,653,513	2,153,140	147,547	6.9%
S2	379,626	1,469,429	1,849,055	120,001	6.5%
S3	285,652	1,292,335	1,577,986	93,974	6.0%
Lower secondary average			**5,580,181**	**361,522**	**6.5%**
S4	207,352	1,080,129	1,287,480	78,300	6.1%
S5	134,326	1,088,489	1,222,815	73,026	6.0%
S6	80,804	1,076,498	1,157,301	53,522	4.6%
Upper secondary average			**3,667,596**	**204,848**	**5.6%**
BD1	58,804	430,404	489,208	21,999	4.5%
BD2	38,050	425,332	463,382	20,754	4.5%
BD3	18,471	419,387	437,858	19,579	4.5%
BD4	-	412,673	412,673	18,471	4.5%
Tertiary average			**1,803,121**	**80,804**	**4.5%**

Table A9.22: Econometric output from returns calculation using SPSS

	Coefficient	t-stat	p-value	Years at level	Implied return
Constant	3.450 (.182)	18.944	.000		
Years of work experience	.037 (.003)	11.065	.000		
Years of work experience squared	.000 (.000)	-.921	.357		
Primary school finished	.092 (.159)	.579	.562	6 years	1.5%
Lower secondary completed	.765 (.172)	4.438	.000	3 years	18.7%
Upper secondary completed	1.680	10.463	.000	3 years	24.2%
Post-secondary education completed	2.352 (.192)	12.233	.000	2 years	29.3%
Tertiary level completed	3.149 (.162)	19.492	.000	4 years	25.4%

Dependent variable = the log of hourly wage for individuals in the sample.
To calculate the final returns estimates required using the inverse power of the number of years for each level applied to.

Sample size

Total = 2074.
Primary school finished = 397 (19.1%); Lower secondary completed = 239 (11.5%); Upper secondary completed = 558 (26.9%); Post-secondary education completed = 109 (5.3%); Tertiary level completed = 585 (28.2%). In addition, no education = 72 (3.5%); Masters completed = 33 (1.6%).

upperquartile