DFID
human development
resource centre

UKaid
from the Department for
International Development

Review of major Results Based Aid (RBA) and Results Based Financing (RBF) schemes

Final report

Mark Pearson, Martin Johnson and Robin Ellison

March 2010

DFID Human Development Resource Centre
HLSP, 10 Fleet Place
London EC4M 7RB

T: +44 (0) 20 7651 0305
F: +44 (0) 20 7651 0310
E: just-ask@dfidhdrc.org
W: www.hlsp.org

Table of Contents

Abbreviations

BMA	British Medical Association
CCT	Conditional Cash Transfers
CODA	Cash on Delivery Aid
CPHF	Centre for Progressive Health Financing
DAC	Development Assistance Committee
DFID	Department for International Development
DHS	Demographic and Health Survey
DRG	Disease Resource Group
DTP	Diphtheria Tetanus Pertussis
EC	European Commission
EC MDGc	European Commission Millennium Development Goal contract
EMIS	Education Management Information System
ES	Expenditure Survey
GAVI	Global Alliance for Vaccines and Immunization
GFATM	Global Fund to Fight AIDS, TB & Malaria
GPOBA	Global Partnership on Output Based Aid
HMN	Health Metrics Network
HRITF	Health Results Innovation Trust Fund
IHSN	International Household Survey Network
IS	Income survey
ISS	Immunisation Services Support
LICUS	Low Income Countries Under Stress
LSMS	Living Standards Measurement Study
M&E	Monitoring and Evaluation
MAPS	Millennium Action Plan for Statistics
MCC/MCA	Millennium Challenge Corporation/ Millennium Challenge Account
MCH	Maternal and Child Health
MIC	Middle Income Country
MICS	Multiple Indicator Cluster Survey
MOU	Memorandum of Understanding
MTEF	Medium Term Expenditure Framework
NAO	National Audit Office
NGO	Non Government Organisation
OBA	Output Based Aid
PbR	Payment by Results
PCT	Primary Care Trust
PRBS	Poverty Reduction Budget Support
PS	Priority Survey
QOF	Quality and Outcomes Framework
RBA	Results Based Approaches
RBF	Results Based Financing
SWAP	Sector Wide Approach
TA	Technical Assitance
UNSD	United Nations Statistical Department

Key messages

"Everything that can be counted does not necessarily count; everything that counts cannot necessarily be counted"
Albert Einstein

"If you can't measure it, you can't manage it"
Peter Drucker

Introduction

"Managing for results" is a key component of the Paris Principles of Aid Effectiveness. The lack of a results focus is seen as a major reason why past aid efforts have yielded disappointing results. Donors are currently looking for ways to scale up support whilst, at the same time, demonstrating the results they are achieving. Results based approaches – whether result based aid (RBA) or results based financing (RBF) – are possible ways of doing this. This report reviews a range of RBA and RBF schemes.

What are RBA and RBF?

RBA and RBF schemes both involve contractual arrangements between a *principal* and an *agent* and involve the transfer of funds in exchange for the delivery of specified results. They differ primarily in terms of their funding sources and who the respective principals and agents are. Some schemes are hybrids of the two approaches.

In order to achieve the desired results the schemes try to align the incentives faced by both the principal and the agent. The schemes reviewed employ a range of performance levers and rely heavily on financial incentives. Alternative approaches are available which use different performance levers; agents are motivated by a range of factors not just financial ones.

RBA/RBF is being employed in situations of complexity and where there is often great uncertainty surrounding, and lack of understanding about, the results chain. This makes assessing results and attribution extremely challenging.

This review raises and addresses a range of questions.

Are we focussing on the right results?

It is important first to take a step back and ask "are we targeting the right results?" The risk is that schemes focus on results that are measurable instead of results that are important – and what gets measured gets done. This issue is important but beyond the scope of this review.

Do RBA/RBF schemes deliver the intended results?

Yes – but that is not necessarily enough. The evidence base for results based approaches is generally weak. Most efforts focus on the question of whether results are *associated* with RBA/RBF than whether they are *due* to them. Whilst schemes generally appear to deliver results (i.e. are associated with better results) it is extremely difficult to say whether this is due to the results focus itself or simply from the additional funding associated with it. There is typically little or no evidence to refute the hypothesis that alternative approaches could not have delivered equally promising or even better results.

Where better results have been found it is still far from clear that these are worthwhile when one considers the costs involved. These comprise both the large, and often unnecessarily large, transaction costs associated with the schemes as well as any negative unintended consequences.

More sound evaluations are needed. Such evaluations need to focus on all of the effects of the schemes – including the broader systems impact – and not just the extent to which they deliver the specified results. They should also consider issues of cost effectiveness. Particular attention needs to be paid to the question of whether the transaction costs are justified by the results achieved, whether there is consistency with the Paris Principles and the extent of unintended effects.

General budget support is a partial exception. The evidence base is relatively strong and generally positive – both in terms of delivering results as well as consistency with the Paris Principles. Whilst efforts are required to make it work better – building on the findings of the various recent evaluations – it should be the instrument of choice and used unless there are compelling reasons not to.

Will attribution be possible?

Generally not. RBA and RBF will not be a simple answer to DFID's attribution concerns. Attribution will continue to be extremely difficult, if not impossible, to assess. As noted above, building the evidence base through well designed evaluations will go some way towards demonstrating whether results are due to the schemes themselves. However, RBA/RBF schemes are rarely, if ever, implemented in isolation but tend, quite rightly, to implemented as part of a broader package. Attributing results to specific interventions – such as RBA or RBF – within such a context adds further challenges.

Are the RBA/RBF schemes consistent with best practice on aid effectiveness?

Generally not. Many of the schemes reviewed run counter to at least some of the other principles of aid effectiveness (notably alignment with country systems and country ownership). This is partly a feature of the institutions which have taken RBA/RBF forward (e.g. GFATM, GAVI which have a disease specific, sub sectoral focus). Those that by-pass government are not aligned, those that involve government may simply be an additional layer of donor interface which add little value. Many of the schemes are narrow which reduces the scope for strengthening the system as a whole and creates risks that they will further fragment the sectors they operate in. DFID needs to be judge whether the benefits achieved from results based approaches (which are often generated fairly quickly) outweigh any costs associated with not adhering to the Paris Principles (which may only be evident in the longer term). In short, a results focus should not be at the expense of other efforts to improve aid effectiveness.

Do RBA/RBF schemes offer value for money?

Unknown. Value for money will need to be carefully monitored. Financial risk can be shifted to agents but this is likely to have cost implications and may potentially be at the expense of service delivery for the poorest (e.g. the most vulnerable countries are likely to respond worst to CODA). This emphasises the importance of aid design in terms of the range of schemes implemented and the need for aid instruments tailored to the requirements of each country context. As noted above schemes also need to be assessed for their cost effectiveness against other alternative approaches.

Will RBA/RBF schemes continue to deliver results over the medium to long term?

Unknown risk. There are questions as to the long term impact of results based approaches with the risk that results will not be sustained. Many of the schemes are in their relative infancy – successes may reflect, in part, the rapid responses of early adopters. Agents may adapt over time as they learn to "game" the system (securing economic rents whilst reducing the risks they bear). There is also a risk that the schemes (RBF in particular) become a "sticking plaster" which may reduce the perceived need for the more fundamental, and often politically sensitive, reforms which might be required to sustain progress. A distinction needs to be made between the valid long term role of RBA/RBF and the shorter term "catch up" role they might play which suggests the need for a long term vision of how schemes will contribute to system development over time.

Does conditionality help?

Sometimes not. Lessons from the evolution of conditionality show that it is most effective when there is a shared commitment to achieving targets (the underlying principle of PRBS conditionality). Conditions continue to be associated with aid more generally, including RBA/RBF. The role of conditionality, however, needs to be continually reviewed. In some (i.e. non budget support) cases, the value added of conditionality is not immediately apparent where there is already a shared commitment to achieving targets. In such circumstances all it may do is signal a lack of trust. Whilst donors may sometimes require conditionality for public relations purposes it should be recognised that this may bring little or nothing in terms of additional results. (There is little evidence, for example. that conditional cash transfers are any more effective than unconditional ones.) In a similar vein, given the existing levers of the conditionality package that is already associated with budget support, is there really a strong case for the additional bureaucracy of variable tranches or do they just compensate for weakness in demonstrating progress and communicating this to taxpayers?

Do RBA/RBF schemes promote equity?

Mixed. This is an issue at a number of levels. In some cases (GAVI ISS, MCA) poorer countries find it hard to access RBA/RBF funding in the first place. This is either because the application process is complex or because a prior performance record is required. There is also some evidence that poorer countries are less able to secure available rewards (GAVI ISS) though this is not always the case (Global Fund). Within countries deprived areas have often performed surprisingly well. Equity is often influenced more by the institution actually implementing the scheme which in some cases dictates the type of services which are targeted that the RBA/RBF approach itself. Targeting of particular beneficiaries is possible but involves costs and significant implementation challenges. This has particular consequences for fragile states where capacity tends to be weakest

Whilst it remains important to ensure equity is emphasised at all stages of the identification and implementation of RBA/RBF schemes there is certainly no evidence to suggest that RBA/RBF schemes will *automatically* disadvantage the poor. A number of features may lend themselves to more equitable results. These might include the use of locally identified targets in low income countries (which might imply lower but still challenging targets as opposed to the use of global targets or standards)...Up front capacity building efforts may be needed to in some cases to help weaker countries take advantage of results based schemes. Different (simpler) approval processes and technical support may also be needed to ensure capacity constrained countries can benefit from RBF in the first place.

To what extent does good design matter?

A lot. Design needs to be informed by a sound understanding of what motivates agents, a good understanding of the results chain and an analysis of the political economy. Good design is essential and can help minimise – but not remove – unintended negative consequences through practices such as gaming and cherry picking and can also help promote equitable results.

What are the key statistical/measurement issues?

Despite careful design, many RBA/RBF schemes still face fraudulent reporting and their narrow focus can work against important broader aspects of progress. Narrow, high-stakes schemes are generally more strongly affected. The established schemes tackle these issues more or less satisfactorily, but some problems may be inherent and cannot be eradicated.

Government administrative systems are often untimely and produce unreliable data. Narrower RBA/RBF schemes often set up parallel administrative systems which are generally more effective, though not perfect. Government household surveys are not conducted very frequently in developing countries and they suffer from sampling errors and biases, which limit their application to wider schemes. These problems with data sources tend to be worse in fragile states.

While the idea of mainstreaming parallel systems into government systems is worth investigating for various schemes, the problems may prove to be greater – and the gains smaller – than might appear to be the case.

DFID and other development partners such as the EC that operate broader schemes need to renew their efforts to encourage and support wider statistical capacity building. The new Statistics for Results Facility may play an important role here.

Conclusion

RBA/RBF schemes certainly have a role to play but are no panacea. As the quotes above suggest – even if we know what counts, this doesn't mean that we can measure it …and if we can't measure it we can't manage it. This review concludes that DFID should adopt a positive but cautious stance in relation to RBA/RBF schemes with a strong emphasis on piloting and rigorous evaluation. Schemes need to be tailored to local circumstances. They should be well prepared, well designed, piloted and carefully monitored and then modified as and when any unexpected effects become apparent. RBF appears to work better for simple interventions which are provider led and where latent capacity exists. Complementary actions will usually be required. Some schemes incorporate a range of approaches – for those that do not, they will either have to be integrated into the existing schemes or coordinated with them.

1. Introduction

For the past few years DFID has increasingly engaged in debate about results based aid and financing (RBA and RBF). This is partly because of the continuous effort to improve the impact of every pound spent in poverty reduction and partly to explore how we can reduce aid conditionality, fiduciary risks and transaction costs while at the same time being more transparent and predictable with its aid.

DFID's 2009 White paper states that: *"The development community has often stood accused of making big investments and bigger promises without taking enough care in ensuring they deliver outcomes on the ground. That situation has changed over the last decade. The UK's approach to increasing the flow of aid through developing country governments to better target assistance, coupled with a focus on aid effectiveness and results, has improved the impact of our aid.[…] the challenges of growth, climate change, conflict, and the rightful concerns of the public require new approaches to delivery. These challenges demand greater efficiency and focus on value for money." (p.125)*[1]

The current interest in results based aid and financing is not surprising given the UK government's commitment to increase development assistance to 0.7% of GDP by 2013, whilst remaining confident that the money is spent on its intended purpose, can be accounted for and provides good value with regard to poverty reduction. Budget support operations have played an important role in recent years, not only in assisting countries finance and implement their poverty reduction strategies, but also in providing a practical vehicle to translate an expanded aid budget into development assistance to alleviate poverty.

Whilst there may scope to use budget support as a means of translating a further acceleration in DFID's aid budget into development assistance that is intended to be delivered within a results framework, there will be constraints in extent to which this can be achieved without affecting the fundamental principles that ensure budget support remains oriented towards poverty reduction, and that issues of fiduciary risk are appropriately managed.

To address the challenges identified in the White Paper and the requirement to scale up aid delivery over coming years, DFID needs to take stock of which elements of its funding are currently conditional on the achievement of specific results. Specifically, which of these could be developed further to ensure that an expansion of development assistance is achieved where it is required, whilst remaining appropriately results oriented and whether alternative and innovative performance based instruments that are being pioneered by development partners (and/or in developed countries) might also have a role to play.

The international debate on the use of results based aid (and results based financing) approaches tend to be polarised, with some aid practitioners strongly in favour and others more sceptical. The evidence of successes and failures of existing results based aid schemes (or the absence thereof) has fuelled this debate. There is a need for greater clarity about what the international community means by results based aid (and results based financing), the extent to which these approaches are being used in practice already, whether such practices should be expanded, the pros and cons and modalities of the different schemes, and how they might be adapted and better used.

[1] DFID, "Eliminating World Poverty: Building our Common Future", July 2009. Square brackets added by author.

The purpose of this paper is to review major results based aid and results based financing schemes funded and co-funded by DFID and by other development agencies and partner countries and to outline how these work in different contexts

The approaches are defined in greater detail in section 3. In short, the key differences between them are in terms of:

- **Funding source** (RBA must be aid funded/RBF can be funded from any source including aid); and
- **Contractual partners** (RBA normally involves a contract between a donor and government – RBF normally involves a contract between government and an implementing partner at sub national level).

Some of the schemes reviewed are hybrids of the two. The schemes reviewed include:

- **Results Based Aid:** GFATM (Global Fund to fight AIDS, Tuberculosis and Malaria), GAVI alliance, Budget Support operations (including PRBS and EC MDG Contracts); Global Programme on Output Based Aid (GPOBA),[2] Cash on Delivery Aid (CODA), Millennium Challenge Account (MCA)
- **Results Based Financing in developing countries:** Health Results Innovation Trust Fund (HRITF) managed by the World Bank, GPOBA[3], voucher and social transfer schemes (including Conditional Cash Transfers); contract based financing and other performance based funding and financing.
- **Results Based Financing in developed countries**: schemes such as the NHS Quality and Outcomes Framework (QOF) in the UK (Payment by Results (PbR) was also included) and Medicare Pay for Performance (P4P) initiatives in the US.

Table 1 (next page), lists each of the schemes reviewed in this paper, categorises each as RBA or RBF (and notes the results terminology employed by the scheme itself) and presents a short description of each.

[2] GPOBA is hybrid RBA and RBF.
[3] GPOBA is both RBA and RBF, as noted in previous comment.

Table 1 – List and short description of RBA / RBF schemes reviewed

Scheme	Category	**Short description** (including DFID funding and status)
GFATM	Results Based Aid/Financing hybrid	Funding for years 3 to 5 dependent on overall performance achieved during first two years of grant implementation. Scheme is established; DFID is providing funds.
GAVI ISS[4]	Results Based Aid	Initial investment based on (self reported) number of children expected to be vaccinated in year 1. Subsequent reward payments of $20 per child vaccinated above this baseline. Scheme is established; DFID is providing funds.
UK QOF	Results Based Financing	Payment made against performance by general practitioners in the UK against over a hundred quality based indicators. Scheme is established. UK based.
UK PbR	Results Based Financing	Set payment (based on national average unit costs) paid to hospitals in the UK for delivering a specific health output (e.g. hernia operation). Scheme is established. UK based.
US P4P	Results Based Financing	Payment made to providers with level based on performance against a range of quality based output indicators. Scheme is established. US based.
Vouchers	Results Based Financing	Reimbursement made to accredited providers on basis of services delivered to voucher recipients. Schemes have been established; DFID is providing funds.
CCT	Results Based Financing (might be usefully categorised as demand side approach)	Payment made to targeted beneficiary in return for them using specified services. Schemes have been established. Heavily focused in middle income countries using domestic funds.
HRITF	Results Based Aid/ Financing hybrid	*Vehicle* for supporting results based financing approaches, The Fund also focuses on raising resources and knowledge generation Schemes have been established. DFID will be providing funds – Norway is main funder.
GPOBA	Results Based Aid/ Financing hybrid (partnership and trust fund)	Multi-donor partnership and trust fund established to (i) fund and facilitate the preparation of OBA projects in which payment is made to an implementing agent – usually private sector but could be NGO and usually in the utilities sector – for each unit of output supplied and (ii) document and disseminate lessons learned. Schemes have been established. DFID is providing funds.
PRBS	Results Based Aid	Payment made to government in return for commitment to good governance and satisfactory progress in poverty reduction. (Variable tranche has some similarity to RBF). Schemes have been established; DFID is providing funds.
EC MDG Contracts	Results Based Aid	Payment made to government in return for commitment to good governance and satisfactory progress in poverty reduction. (Variable tranche has some similarity to RBF.) Schemes have been established but are new; DFID will be providing funds (indirectly through EC contribution) with potential to supplement with bilateral funds
CODA	Results Based Aid (progress based aid)	A concept for making payments to government in return for achievement of specific results (e.g. increase in primary school enrolment). Yet to be established.
MCA	Results Based Aid	Payment made to government in return for demonstrable commitment to democracy, good governance, 'economic freedom' and pro-poor public services. Scheme has been established; no DFID funding. US funded

[4] The review focused on GAVI ISS. GAVI also has an HSS window though it only started recently and evidence is just beginning to emerge.

This work supports implementation of DFID's overall Results Action Plan (published in January 2008), the key priorities of which are shown in Box 1. These are in line with the "Managing for Results" theme set out in the Paris Principles which emphasises the need to focus on dialogue on results at all phases of the development process, align programming, monitoring and evaluation with results, keep measurement and reporting simple, managing for, not by, results and using results information for learning and decision making (Marrakech 2004).

Box 1: DFID's Results Action Plan – The Ten Priorities

WITHIN DFID

- More use of quantitative information to improve decision-making;
- Further strengthened performance and results frameworks for country programmes;
- Improved communication to the UK public on the results of development assistance;
- Review of people management systems to encourage a stronger focus on outcomes;
- Independent Advisory Committee for Development Impact established to strengthen the independence of the evaluation function.

WITH PARTNER COUNTRIES

- Investment in statistics through internationally coordinated funding;
- Support accountability mechanisms to scrutinise governments' and donor performance.

INTERNATIONALLY

- Support an internationally coherent approach to impact evaluation;
- Promote new international mechanisms for mutual accountability between donors and partners, and seek agreement at the Ghana High Level Forum in 2008;
- Promote new international mechanisms for assessing agency effectiveness, and seek international agreement at the Ghana High Level Forum in 2008.

The review is structured as follows:

- **Section 2** presents a framework for analysis based on a theoretical overview of the issues
- **Section 3** attempts to make some sense of the current confusion related to the use of terminology and provide a suggested working definition of RBA/RBF:
- **Section 4** describes the current approaches adopted by a variety of schemes according to dimensions set out in a template presented in annex 1
- **Section 5** looks at statistical and measurement issues
- **Section 6** provides an analysis of the schemes assessing what is know in terms of impact and outlines emerging lessons at all stages of identification, design and implementation
- **Section 7** sets out conclusions and possible next steps for DFID

The completed templates are presented in a separate annex.

2. Framework for analysis

This section very briefly illustrates one of the key challenges faced by RBA/RBF approaches – the issue of how to deal with the principal-agent problem. It highlights the fact that whilst financial incentives – a key component of RBA/RBF – are important they are only one of a number of incentives which agents face. It also shows that RBA/RBF approaches use a number of performance levers but these only represent a subset of the overall options available to policy makers.

2.1 The principal agent problem

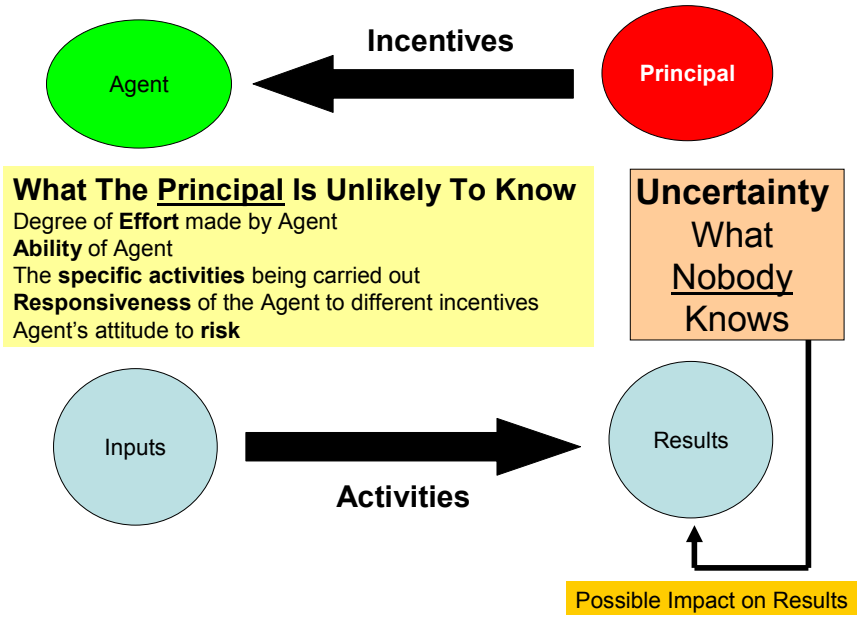
The principal agent problem refers to a situation in which a principal (who wants to achieve certain results) tries to ensure that its agent (who actually undertakes the activities needed) delivers the desired results. Problems arise where there is incomplete or asymmetric information which the agent can take advantage of to pursue his or her own agenda. The challenge facing the principal is to establish a mechanism – usually through some form of contract – in which the incentives faced by the two parties are aligned. A key issue in any contract is the extent to which risks are transferred from principal to agent. Under more traditional arrangements – where payment does not depend upon results – the risk is borne by the principal. RBA and RBF schemes are designed to shift *some*, and some cases *all*, of the risks to the agent. However, the agent will expect to be paid more to compensate for the additional risks they face.

The issue is particularly acute within the aid environment as asymmetries of information exist in a variety of forms:

- Between donors and government (which is relevant in terms of results based aid),
- Between governments and services providers (which raises further issues in terms of results based financing approaches) and
- Between donors and service providers (where the donor engages directly with the service provider on behalf of the government – as with schemes funded under GPOBA).

The issue is illustrated in chart 1 below, which outlines some of the main areas where the principal may be lacking knowledge or information.

Chart 1: The Principal Agent Problem

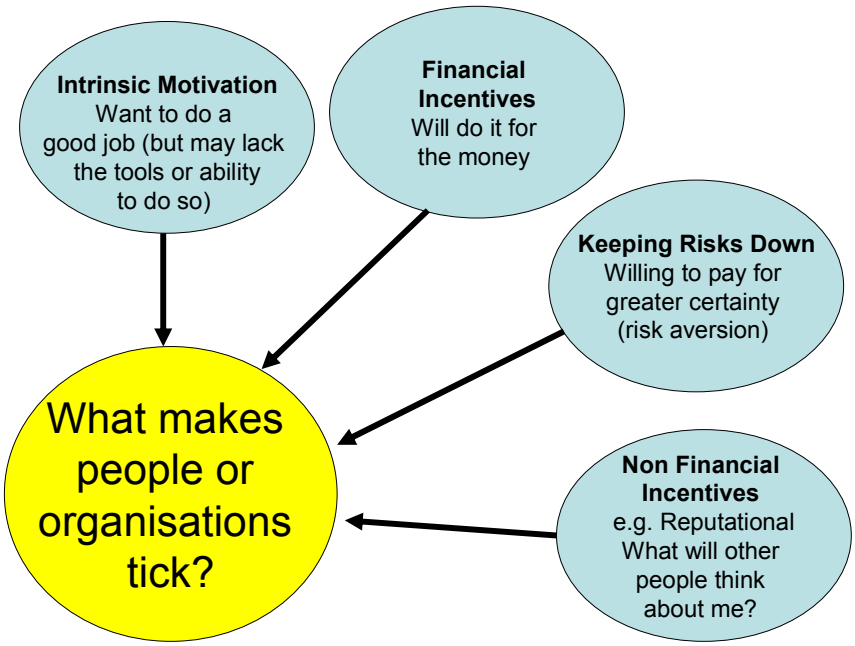


The problems of agreeing a contract are further increased by the level of uncertainty associated with the achievement of any results. Some of this uncertainty can be addressed, at least in the medium term, by policy action. For example, building a strong evidence base can help reduce any uncertainties in the results chain. However, some uncertainty is purely external. The current global financial turmoil, for example, is having far reaching effects in low income counties. This will affect countries' ability to provide, and the population's ability to access, quality health care in ways which could not have been predicted two or three years ago.

2.2 What motivates agents?

If we are concerned that agents will pursue their own individual agendas it is important to know what their underlying motives might be. An immediate reaction is often to assume that if you pay somebody enough to do something they will do it. There may be *some* truth in this but agents are also motivated by other factors as shown in **chart 2** below. People might simply want to do the right thing but may lack the skills or tools to do so. They might like to avoid risk (an implication being that if they are required to take on more risk, as implied by RBA and RBF, this would need to be reflected in a higher level of compensation). Finally, there are other non-financial motivations such as reputational factors – people really do care what others think about them and their performance. Given that we rarely, if ever, know what motives agents have it becomes extremely difficult to predict their responses to any incentives they are presented with.

Chart 2: What motivates agents?



For example, Mellstrom et al[5] found that when financial rewards were introduced for blood donations in Sweden the overall number of donors declined. This was because most donors were motivated by social or moral reasons and stopped donating when the process became "commercialised". Similarly, a day care centre in Israel found that the introduction of a small financial penalty for parents arriving late to collect their children resulted in an increase the number of late arrivals (Levitt and Dubner, 2005) as it removed the social disapproval previously associated with such behaviour.

The agents' response will depend on their individual motivations as well as the range and level of incentives they face. Larger financial incentives are likely to achieve a larger response. The key question is how to design a range of incentives and set them at a level which achieves the behaviour change required to achieve the desired results but do so at the lowest cost (why pay $10 when they would do it for $5?).

One hypothesis might be that with many agents in low income countries earning little personal income – and sometimes even at or below subsistence level – financial motivations might be stronger than in developed countries settings, though this would need to be demonstrated.[6]

The key implication of this is that in trying to encourage agents to undertake certain tasks it is important to have a clear idea of their underlying motivations. It also raises the question of whether DFID might make more use of non financial incentives. Should DFID publicly acknowledge a Minister for Health who has performed exceptionally well and has demonstrated significant commitment to reducing poverty in his/her country by putting a profile of their website? Would this be more effective that providing additional funds to the ministry?

[5] "The supply of blood donors decreases by almost half when a monetary payment is introduced".
[6] This observation would not necessarily apply to all innovative RBA or RBF schemes, however, particularly where the formal private sector is contracted as the agent, as is usually the case with GPOBA-type schemes.

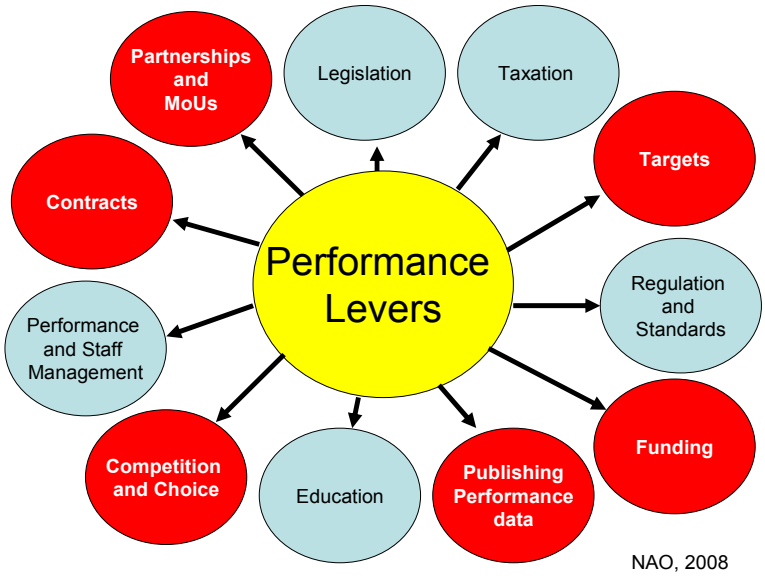
2.3 What are the available performance levers?

Having understood the underlying problem and acquired a good knowledge of what factors motivates the agent, the next question relates to the choice of levers which are available to promote the desired performance.

Chart 3 (based on NAO 2008) outlines a range of potential performance levers. These range from macro-level, market based levers such as taxation, through administrative or bureaucratic levers such as legislation to more micro-level levers such as contracts and right down to the individual level in terms of staff management.

The chart highlights those levers which are particularly relevant for RBA/RBF approaches (in red/darker shading). All the schemes reviewed here have funding attached (although not all of the incentives provided are financial). Some involve targets – others identify the results in other ways (e.g. the *number* of services provided or the *quality* of services delivered). Very few involve publishing performance data (which would tend to try and play on agents' concerns about reputational risk[7]). All of the approaches use some form of contract or Memorandum of Understanding (MoU) though these vary enormously in scope from legally binding contracts to an implied *social* contract (i.e. that recipients of conditional cash transfers will spend their money wisely). Some involve competition and choice though this is applied in different ways. Under the Millennium Challenge Account countries compete to access funds according to ex ante conditionality. Under Payment by Results in the UK providers are supposed to compete on quality (given that prices are fixed) and in some US P4P schemes reward payments are assessed on the basis of performance in relation to other providers (i.e. reimbursement for those providers who perform worst on quality is lower than for those who perform well).

Chart 3: What are the available performance levers?



NAO, 2008

[7] Under QOF data are published though readers are cautioned not to use it as a basis for comparing the relative performance of different providers. This begs the question of what it actually does do.

2.4 What is performance? What are results?

Identifying levers that can influence performance is one thing but this is rather pointless without deciding first which results we want to achieve. Relevant Development Assistance Committee (DAC) definitions are presented in box 2 below.

Essentially "results" can be whatever we want them to be – it depends on the context. They are typically defined by level – output, outcome or impact[8] – but often focus on particular dimensions within these levels. Donors, for example, are particularly interested in equity so often want to know who is using services and who is benefiting from them. In developed countries the focus has often been on quality (people are using enough services – it's just that they are not good enough). The next two sections show how this might be applied in the health and education sectors.

Box 2: Key Definitions

Result: The output, outcome or impact (intended or unintended, positive and/or negative) of a development intervention. Related terms: outcome, effect, impacts
Performance: The degree to which a development intervention or a development partner operates according to specific criteria/standards/ guidelines or achieves results in accordance with stated goals or plans.

(Source: DAC)

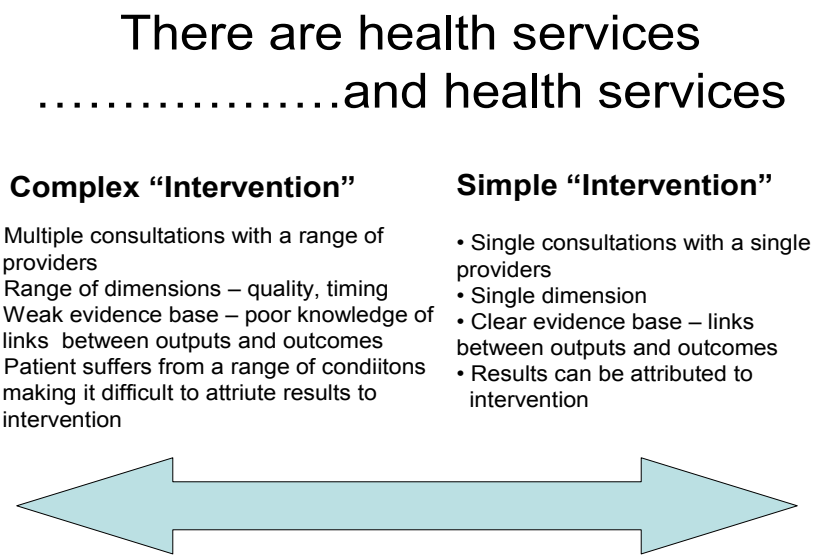
2.5 Example of context for RBA/RBF: the health sector

Health care involves complex processes which often make it extremely difficult to know whether good performance has been achieved and, if it has, who or what should get credit for it. Quality, in particular, is a complex and subjective concept (quality from whose perspective, patients or doctors?[9]). Equity is also particularly difficult to measure as it can be viewed from a number of perspectives. As shown in **chart 4** there will be a spectrum of services ranging from:

- A **simple intervention** such as immunisation where the intervention is delivered by one provider often at one sitting (some vaccinations do require multiple shots), where quality is less of an issue (provided the vaccine has not passed its expiry date and any cold chain requirements have been met) and where the link between the intervention and health outcome is well known.
- A **complex intervention** involving multiple interactions with a range of different health providers for a patient suffering from a range of pre existing conditions. In such cases, it is almost impossible to know whether the intervention actually worked and if it did, why. A key problem is that providing incentives for good performance for a few specific elements of a single disease may lead to neglect of other, potentially more important elements of care.

[8] In section 3 we also argue they can also be set in terms of inputs.
[9] Harold Shipman, a notorious GP responsible for murdering numerous of his patients was extremely popular with his patients http://www.trutv.com/library/crime/serial_killers/notorious/shipman/dead_1.html

Chart 4: the importance of context: health



2.6 Example of context for RBA/RBF: the education sector

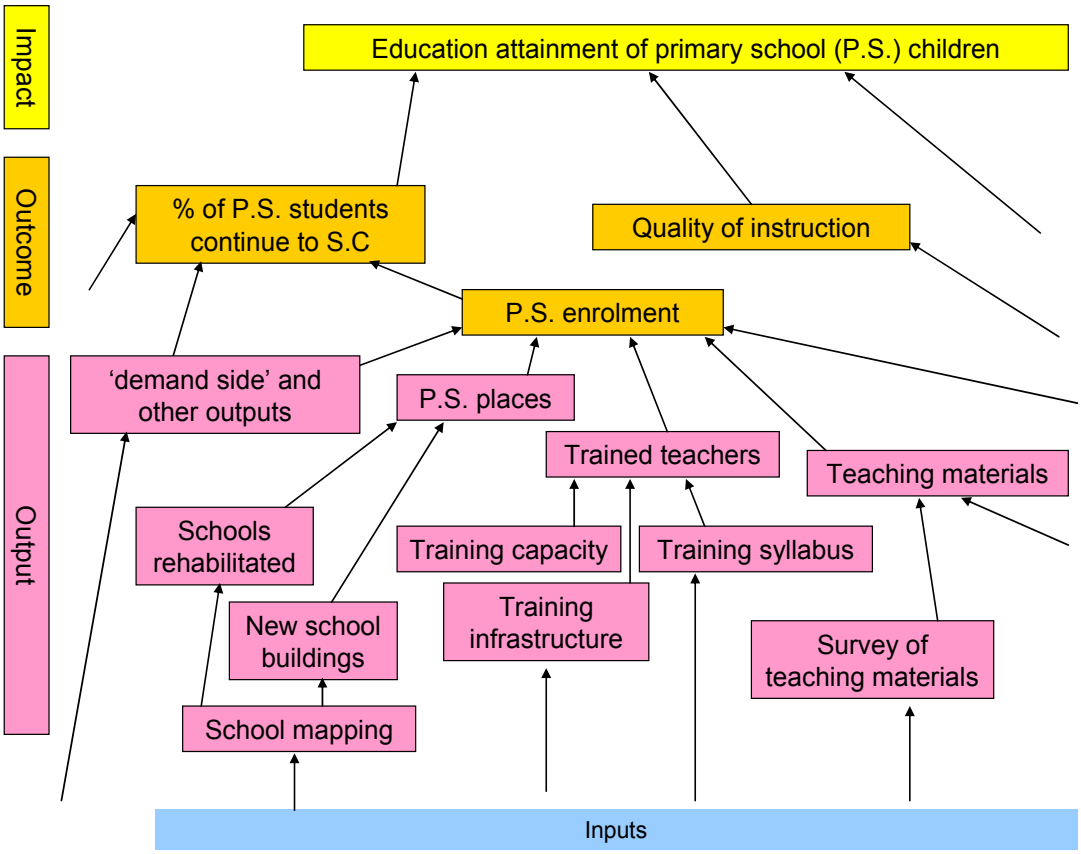
Applying a results focus to initiatives in the education sector is equally problematic. There are few, if any of the 'simple interventions' such as those described above. Initiatives are beginning to focus less on simple enrolment and more on retaining students in the education system until they at least complete primary education – typically lasting six years or so – and on improving the quality of education they receive.

In contrast to the heath sector, indicators and data sources on outcomes are much less satisfactory. The measurement of learning outcomes and linking these to specific interventions is much less straightforward than for outcomes such as changes in mortality and morbidity in diseases (e.g. as a result of immunisation or use of bed nets). National progress often proves hard to identify, and attributing any progress to individual initiatives often impossible. 'Stability in testing' educational achievement is notoriously difficult. This is true even in countries like the UK, which has invested huge financial and human resources in recent years.

A complex mix of interventions (school buildings, teaching and learning materials, curriculum, nutrition, teachers, school management and supervision, community involvement, etc.) is required to retain children in the system over many years, to create a good educational environment and to ensure a given level of quality. The research on which factors are most effective and in what combination is far from clear. Chart 5 provides an indication of the complexity of the results chain.[10]

[10] The purpose of chart 5 is to *illustrate* complexity rather than *identify* all of the factors which affect educational outcomes There are a number of arrows feeding in from unidentified outputs and outcomes at all levels to demonstrate that lots of other factors (including factors outside of the education sector, e.g. domestic violence, nutrition, etc.) which influence outcomes.

Chart 5: Complexity of the Results Chain: Education



See also annex 2.

Last, there is a lack of clear leadership in developing education statistics. There is no corresponding organisation to the Health Metrics Network that might be able to develop thinking on the best indicators, data sources, techniques and national information strategies that might combat some of these problems.

Key messages:

- There are inherent problems in ensuring good performance when those responsible for delivering the performance are not those setting the goals or objectives
- It is important to understand what motivates the individual agents
- Over reliance on financial incentives can lead to undesirable results – it's not just about the money
- There are a range of levers which might be used to align interests. RBA and RBF use some of them
- Results can be whatever you want them to be – it depends on context. But if you focus on one thing – it may lead to the neglect of others
- The complexity of the social sectors makes the issue of choosing a point in the results chain (for a desired impact), measuring results and attributing causality particularly challenging.

3. What Is RBA/RBF?

3.1 Introduction

This section describes the difference between the various results based aid and results based financing schemes: It uses the standard DAC definitions (shown in Annex 3) with the following caveats and comments:

- **Outputs are not homogenous**: the DAC definitions are fairly clear cut with a single definition covering *all* outputs. In practice, there are major differences in the extent to which outputs link to outcomes in the results chain. In some cases the output will actually be an excellent proxy for an outcome. (For example, if a child is vaccinated there is very little that can prevent the outcome being achieved). For another output, e.g. skilled birth attendance, there are major quality issues and many factors can prevent an outcome being achieved). There can also be a hierarchy of outputs. Some would see the production of trained health workers by a training institution as an output in itself. If so the links to outcomes are extremely weak. Will the health worker use the skills learnt? In the public sector? Will he or she migrate? Others would see the trained health worker as an input – or perhaps an intermediate output? What is clear is that different outputs will play different roles in any results chain and the strengths of their links with higher levels such as outcomes may vary considerably.

- **By definition outputs are what is being delivered**. This leads to problems in comparing across programmes and sectors. In education, school enrolment (utilisation of a service) is often seen as an outcome. In health, immunisation is often seen as an output even though it also represents utilisation of a service and is arguably a far better proxy for impact (as noted above – you immunise kids they are protected).

- **Relationships within and between various levels of the results chain are not linear and are interlinked** (see Chart 5 and associated text above).

- **The DAC definition of results needs to be expanded to include inputs for the purpose of analysing RBA and RBF schemes.** Some budget support operations include inputs as performance indicators (e.g. a floor under budget allocations and/or expenditure in particular sectors or sub-sectors such as health and education).

It should be noted that individual **RBA and RBF schemes should not be seen in isolation.** Many form part of a broader package which might incorporate other forms of financial assistance, including up front investments, to enable countries to develop the capacity that is required to take advantage of any financial incentives on offer. Equally, up-front investments and financial rewards for performance are often complemented by other forms of support such as technical assistance and policy dialogue. Any assessment of effectiveness for individual RBA and RBF schemes needs to be carried out in this light.

Key messages:

- a good understanding of the results chain is essential
- need to be clear that ‘outputs’ are what is to be delivered
- some outputs and outcomes exert much greater influence on ‘impact’ than others
- there are differences between sectors in terms of relative proximity of outputs and outcomes to impacts

- need to consider RBA and RBF in the broader context of what they are trying to achieve – they are rarely applied in isolation but are often part of packages.

3.2 Distinguishing between RBA and RBF

Working definitions of RBA and RBF are presented in box 3 below. They are based on the view that whilst the different schemes may adopt different approaches to setting up contracts between principal and agents and may use different ways of rewarding good performance they do share key features – **they do use contracts**, **they do identify the desired results** and **they do provide funding which is linked in some way to performance** against these results (however implicitly this is done and even if this is not the main objective).

Box 3: Working Definitions of RBA and RBF

Results Based Aid. Delivery of aid direct to government through a contractual arrangement that specifies results to be achieved in return for payment to be made.

Results Based Financing. Use of government resources, in a contractual arrangement between government and implementing agent, (sub-national government or non-government) that specifies results to be achieved in return for payment to be made.

Hybrid Results Based Aid/Financing. Use of donor resources, in a contractual arrangement between donor and non government implementing agent, that specifies results to be achieved in return for payment to be made.

Where they differ in a <u>fundamental</u> sense is in terms of their funding sources and institutional arrangements

- **Funding source:** Some of the schemes reviewed are aid funded (e.g. GAVI ISS), some are Government funded (e.g. PbR in the UK) and some may be a combination of the two (e.g. vouchers and conditional cash transfers). **By definition, results based aid <u>must</u> be aid funded** – they cannot be government funded. However, **aid may also be used to finance or part finance results based financing initiatives** (e.g. schemes such as those funded through GPOBA) and such schemes, as a result, can be considered as hybrid RBA/RBF. Pure RBF schemes normally involve domestic funding (the main exception being donor-funded RBF schemes that are hybrid RBA/RBF as noted above).[11]
- **Institutional arrangements – where the funds go (and who is responsible for delivering results)**. **With RBA, a contract is established between the donor and national government**, funds flow from the donor to the national government and the national government is responsible for delivering the results (although in practice it may delegate responsibility for doing part or all of this to others). **With RBF a contract is normally established between the national Government and an implementing partner** (either a sub-national government entity, a private sector organisation or an NGO), funds flow from the government to the implementing partner and the implementing partner is responsible for delivering results.

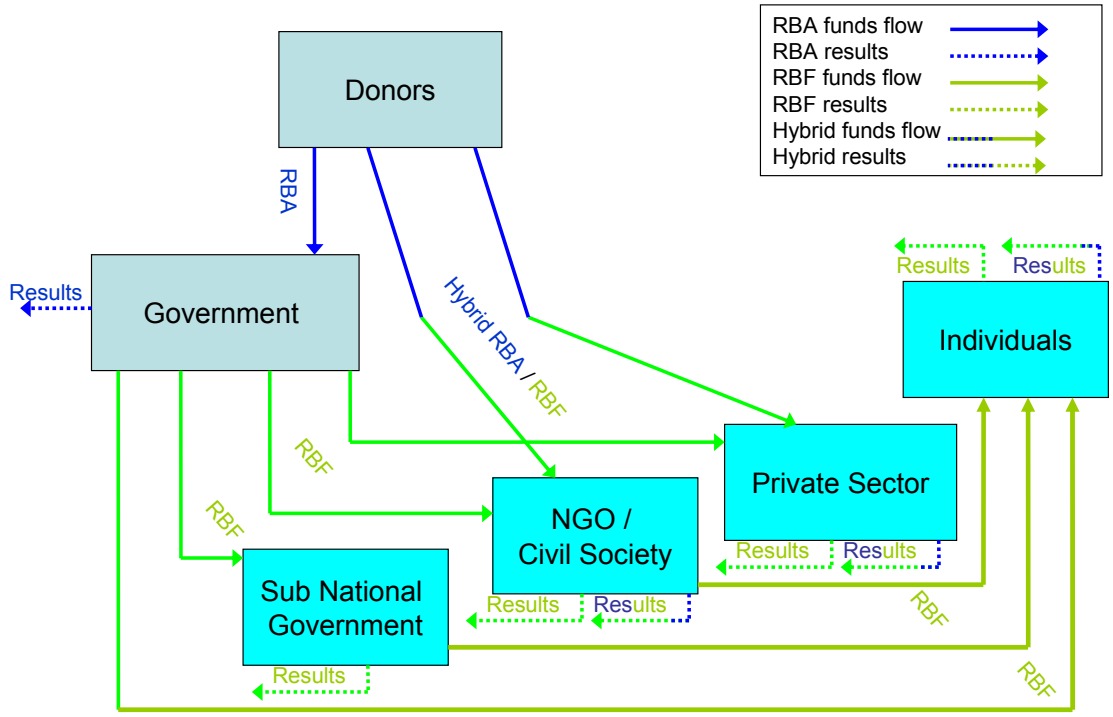
With hybrid RBA/RBF, a contract is established between a donor and an implementing partner (either a private sector organisation or an NGO), funds flow from

[11] Although, in principle, other actors such as NGOs could also establish contracts with an implementing partner and provide finance for RBF-type schemes.

the government to the implementing partner and the implementing partner is responsible for delivering results. With some hybrid RBA/RBF schemes, multiple contracts might be needed. In the case of conditional cash transfers, for example, a donor might contract with an NGO to manage a conditional cash transfer programme. Here the NGO is an agent for one contract and a principal for another.

Chart 6 illustrates these relationships by showing the flow funds between the principal and agent for each type of scheme and where delivery of results takes place for each type of scheme. In each case, a contract is established between principal and agent. In the case where 'individual' is the 'agent' (i.e. for conditional cash transfers) the contractual relationship is in the nature of a 'social contract' rather than a formal contract.

Chart 6: RBA and RBF funds flow (who pays and who receives?)



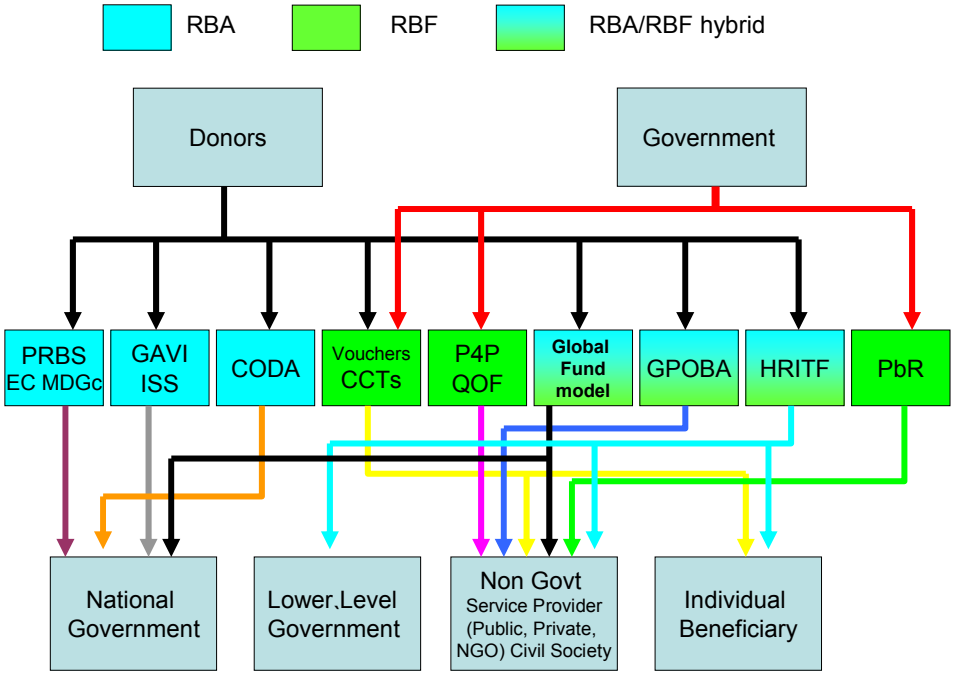
An implicit assumption is that RBF schemes employ the right implementing partner i.e. the one best able to deliver the defined results.

Applying such an approach to the various schemes would suggest that general budget support provided through PRBS and EC MDG Contracts, Cash on Delivery GAVI ISS, GFATM grants and grants provided under the Millennium Challenge Corporation (where contracts are established between donors and government and donor funds flow direct to government) can be classed as RBA. Hybrid RBA/RBF schemes would include schemes such as GPOBA where the contract and funds flow is between donor and non-government implementing partner. GFATM could also be termed a hybrid RBA/RBF scheme when the contract (and funds flow) is between the donor and a non-government implementing partner[12] (known as the principal recipient) but is a pure RBA scheme when the contract (and funds flow) is between the donor and a government. HRITF is another hybrid – all funds come from donors yet all agents are sub national government or other non government agencies (though schemes are designed with support of government and funds may be channelled through government systems). All of the others, with the

[12] This is the case for around half of all its grants (GFATM website).

exception of hybrid schemes, are RBF. Analysis of the flow of funds in the various schemes is shown in chart 7.

Chart 7: Analysis by scheme of who pays and who receives



Degree of country ownership

Where donors contract with an implementing partner other than government a key distinction relates to whether the donor is operating on behalf or government or is doing so to deliberately bypass government which obviously has major implications for country ownership. In practice, the distinction is often not clear cut. In the case of GPOBA, for example the Ministry of Finance has to give a "no objection". In one sense, hybrid schemes are analogous to an RBA contract being established between a donor and a government whereby funds are transferred with the agreement that the government will establish an RBF contract with a non-government organisation for the purpose of delivering the desired results. Depending on the terms of the contract, such an arrangement would be more consistent with the Paris Principles for aligning aid with government systems but would probably raise transaction costs. A key question for hybrid schemes would be whether the donor, in agreeing a contract with an implementing partner, is acting as an effective agent on behalf of government (i.e. doing something Government would have done anyway but reducing transactions costs by going direct to the implementer).

Where the contract is between donor and implementer because of concerns that funds may be diverted to other uses, the donor may *not* be acting as an effective agent of the government. This raises important issues related to aid effectiveness in the use of hybrid schemes which are discussed in more detail later in the document.

Use of terminology

Problems of definition are compounded by the often misleading terminology used by the different schemes. This arises from the use of different terms to represent the same thing, or very similar concepts, and the use of some terms that sometimes represent more than one concept.

With regard to funding issues, the source is made explicit in some cases (i.e. "aid" clearly means external donors) but not in others ("financing" can be taken to mean any source). However, just because something has aid in its title does not mean it is RBA. Although "output based aid" is aid funded, the implementer is not actually government – it is a hybrid scheme rather than pure RBA. For other schemes, different terms such as "pay" and "cash" are used. Pay should be considered to be similar to "financing" (i.e. it can involve either government or donor funding). "Cash", however, is used differently in different settings. "Cash" on Delivery Aid refers to aid money. Conditional "cash" transfers, however, can be aid funded or domestically funded. "Quality and outcomes framework" does not refer explicitly to funding but implicitly means domestic financing in the UK context.

There is probably little point DFID entering into a global debate on definitions of RBA and RBF. This would take time and resources to possibly little effect as donors are probably quite attached to the current terminology. However, given the confusion this causes and likely future interest in these approaches it may well be worth DFID considering a redefinition for internal purposes. ***Our suggestion would be that in order to avoid ambiguity, schemes should be defined according to:***

- ***Their funding sources;***
- ***The contracting parties and the implementing agency responsible for delivering results;***
- ***The types of results desired; and***
- ***The links between results and payments.***

Key messages:

- RBA must be aid financed. The contract is between the donor and government and government is responsible for ensuring results are delivered.
- RBF is supported through government funds, contracts are between government and a sub-national or non-government implementing partner. Central government is not directly responsible for the delivery of results.
- Hybrid RBA / RBF schemes can be supported through donor and/or domestic funds, contracts are between donor and a non-government implementing partner. Central government provides some degree of approval but is not party to the contract and not responsible for delivery of results.
- If DFID plans to pursue approaches in this area it may be helpful to develop its own internal definitions.

4. Overview of schemes reviewed

4.1 Brief overview of schemes

Tables 2 and 3 present a brief overview of the schemes reviewed. A more detailed assessment of the performance levers used by the schemes and the key design features are shown in annex 4 and annex 5 respectively.

Table 2: Overview of key features of Results Based Aid schemes (with additional explanatory information)

	Definition	Recipient of Finance	Results Targeted	Payment Rule	Comments
Poverty Reduction Budget Support (PRBS)	A package agreed between DFID and partner government comprising transfer of un-earmarked budget support resources plus arrangements for conditionality, dialogue, technical assistance, harmonisation and alignment oriented towards the achievement of national poverty reduction policies.	Government	A range of results, from several levels of the results chain (inputs, outputs, outcomes) agreed between DFID and partner government and related to achievement of national poverty reduction policies	Fixed tranche depends on 'satisfactory progress' in implementing national poverty reduction strategy and continued observance of 3 Partnership principles[1]. Does not necessarily require specific 'results' to be achieved as agreed. Variable tranche depends on delivery of small number of specific results from national poverty reduction strategy.	'Satisfactory progress' involves judgement, taking into account results achieved and impact in practice of risk assumptions underlying framework for agreed results. Performance tranche to be disbursed in year 'n+1' depends on progress against specific 'results' in year 'n-1'
EC MDC Contracts	A package agreed between the European Commission (EC) and partner government comprising transfer of un-earmarked budget support resources plus arrangements for conditionality, dialogue, technical assistance, harmonisation and alignment oriented towards progress in eligibility criteria (implementation of national strategy, macroeconomic stability and PFM improvement).	Government	A range of results, from several levels of the results chain (inputs, outputs, outcomes) agreed between EC and partner government and related to progress in eligibility criteria (implementation of national strategy, macroeconomic stability and PFM improvement). Associated indicators are normally presented in a harmonised (multi-donor) performance assessment framework	Fixed tranche depends on 'positive trajectory of change' with regard to eligibility criteria (implementation of national strategy, macroeconomic stability and PFM improvement). Does not necessarily require specific 'results' to be achieved as agreed. Variable tranche depends on delivery of number of specific results from national poverty reduction strategy.	'Positive trajectory of change' involves judgement, taking into account results achieved and impact in practice of risk assumptions underlying framework for agreed results. Variable performance component of up to 30% with two elements:(a) MDG-based tranche: at least 15% of the total commitment would be used specifically to reward performance against MDG-related outcome indicators and (b) Annual Performance Tranche: In case of specific and significant concerns about performance with respect to the implementation Poverty Reduction Strategy Papers (PRSP), progress with PFM improvements and macroeconomic stabilization, up to

					15% of the annual allocation could be withheld..
Cash on Delivery Aid	Contract established between donor and government specifying level of payment to be made by donor for each unit of result delivered as specified in the contract	Government	Potentially any level, although likely to be 'output' or 'outcome'. Current intention is to pilot the concept by targeting primary school completion.	Payment would be proportionate to number of qualifying 'results' delivered.	Currently a concept rather than an active RBA scheme. May involve individual contracts with participating governments or an 'open contract' that any (pre-qualifying) government may sign up to.
GAVI ISS	Initial investment based on (self reported) number of children expected to be vaccinated in year 1. Subsequent reward payments of $20 per child vaccinated above this baseline	Government	Outputs: Immunisation coverage	Payment would be proportionate to number of qualifying 'results' delivered (immunisations above baseline)	
GFATM (Hybrid RBA/RBF)	Proposal based Funding for years 3 to 5 dependant on overall performance achieved during first two years of grant implementation	Government or Non Government e.g. Private Sector, Civil Society	Range of country set targets; contextual factors are also considered. Results focused on the three diseases	Determined through Board decision	Reward takes the form of continued payment. Good performers are also eligible for access to the Rolling Continuation Channel which promises follow on funding with lower transaction costs
Output Based Aid (incl. schemes funded through GPOBA) (Hybrid RBA/RBF)	Contract established between donor and implementing agent which specifies result(s) to be delivered by implementer and level of payment for each unit delivered.	Implementer – usually private sector org., but could also be a public utility, NGO, community-based organization, or a well-run publicly owned company	Outputs	Payment is proportionate to number of qualifying outputs delivered.	Contract usually specifies one or very few outputs (e.g. …) Also a RBA scheme
Millennium Challenge Account	Payment made to government in return for demonstrable commitment to	Government	*A priori* achievement against 17 indicators associated with democracy, good governance,	Once a compact is signed, the whole funding for the life of the compact is committed.	Countries compete each year for access to limited compact funds.

	democracy, good governance, 'economic freedom' and pro-poor public services		'economic freedom' and pro-poor public services.	Resources are released according to a quarterly rolling annual disbursement forecast that is developed by and agreed with the 'accountable entity' (i.e. the agency responsible for managing the funds in-country). Countries that just fail to achieve eligibility for a compact may qualify for Threshold Programme Assistance (smaller amounts of money which target improvement in one or more indicators to improve chances for compact eligibility)	Partner countries propose own "accountable entity" structure to manage and implement compact. 'Threshold' funds target improvement in one or more indicators (to improve chances for compact eligibility).

Table 3: Overview of key features of Results Based Financing schemes (with additional explanatory information)

	Definition	Recipient of Finance	Results Targeted	Payment Rule	Comments
Pay for performance (P4P)	Set payment made against performance of heath providers against a range of quality based output indicators	Implementer	Could be any level (In practice the US P4P focuses on quality of service outputs	Payment is proportionate to number of qualifying outputs delivered and related performance of other agents	In principle very broad. In practice, narrow results focus
Conditional cash transfers	Payment made to targeted beneficiary on the basis of utilising specified services	Beneficiary	Could be any level. (In practice, usually related to utilisation of service outputs	All or nothing	
Vouchers	Reimbursement made to accredited provider on basis of service delivered to targeted beneficiary	Implementer	Could be any level. (In practice, usually related to utilisation of service outputs	Directly proportionate to the number of services delivered by the health provider (which is, in turn, based on the number of vouchers gained)	
Payment by results	Set payment (based on national average unit costs) for services delivered	Implementer	Could be any level. (In practice – relates to the quantity component of service outputs	Directly proportionate to quantity of specified services delivered	In principle: Could use any funding source(As applied to date in UK uses Government funds)
Quality and outcomes framework	Set payment made against performance of GPs against ~ 135 quality based indicators	Implementer	Could be any level. (In practice – relates to the quality component of service outputs	Proportionate to performance against specified quality targets delivered	

4.2 Major differences between schemes

Firstly, there is a fundamental difference between approaches which are actually **schemes** (e.g. individual budget support operations and individual voucher programmes) and those which are **vehicles** for a range of possible results based approaches. The latter may implement a range of schemes and may also have broader objectives beyond simply implementing schemes. In the case of the Health Results Innovative Trust Fund, for example, the objectives are to help build the knowledge base and also to investigate the potential to leverage additional funding for health as well as to implement individual RBF schemes which it does through a variety of demand and supply side approaches. In the case of GPOBA, a multi-donor partnership and trust fund established in January 2003 by DFID with the World Bank, this was originally established to *facilitate* the preparation of Output Based Aid (OBA) schemes, document and disseminate lessons learned and advocate mainstreaming of OBA approaches. In 2005, GPOBA developed its own facility to directly fund OBA schemes in developing countries, which it uses to pilot and test OBA approaches.

Secondly, there is a fundamental difference between **results based aid** and **results based financing** (as discussed above in section 3).

Finally, there are a whole range of more **specific design features** which differentiate the various approaches. Within this category it is possible to distinguish between the broad approaches (the types of levers used) and the real "nuts and bolts" of the approaches. In terms of levers (refer back to **chart 3**) all of the approaches involve some form of funding but vary widely in their use of the other levers. An overview of the performance levers used by the different schemes is presented at annex 4. Some of the main design features are reviewed in the following paragraphs.

4.3 Design features which differentiate the various schemes

The following design features have major implications for the expected results, the effectiveness of the schemes and their consistency with the Paris Principles on Aid Effectiveness.

How the term "results" is interpreted and the types of results used. Again terminology serves to confuse matters. The various schemes use a number of terms, including **results, performance and progress,** interchangeably. Often they mean the same thing, though they tend to be used in different settings. All schemes set out their expected results or, at the least, their broad expectations. In the case of budget support and EC MDG Contacts, for example, results can be quite broad – such as poverty reduction, human rights and good governance. The interpretation of achievement against these broad categories is often informed by more specific targets that are articulated in multi-donor performance assessment matrices that form an important component of the 'budget support dialogue' with partner governments. At the other extreme, results can be simple and easily quantifiable. Under PbR, for example, a result might be the completion of a hernia operation. Similarly, OBA results would include things like connection of a household to a water supply network, or a gas supply network.

The next operational issue is how to assess progress towards, or performance against, these desired results or targets.[13] The term "progress" tends to be used where

[13] Another possible confusion relates to the fact that performance suggests relative progress against a fixed target or goal whilst progress gives more of a sense of direction or absolute change less closely tied to an

assessments are relatively complex and/or involve subjective judgements involving a wide range of indicators at many levels. This is the case, for example, with regard to assessing the results of budget support under PRBS or EC MDG Contracts where variable progress may have been made against a number of pro-poor indicators (including, in some cases, no progress at all). DFID, in fact, refers to 'satisfactory progress' in this regard, whilst the EC employs a 'dynamic interpretation' of its eligibility criteria when assessing progress. The term "performance" tends to be used when it is easy to quantify the degree to which results have been achieved. Thus, although they involve different approaches, assessing progress and assessing performance can be treated as essentially the same thing (with the latter operating with a greater degree of observable precision than the latter).

Some of the schemes do not refer explicitly to results, performance or progress in their title. Cash on Delivery does not mention results but can be interpreted as meaning cash on delivery of results. The envisaged results are things like school enrolments (which is usually treated as an *outcome* in the education sector despite being relatively less close to the desired impact than would be the case for some *outputs* in other sectors – e.g. health).

Chart 8

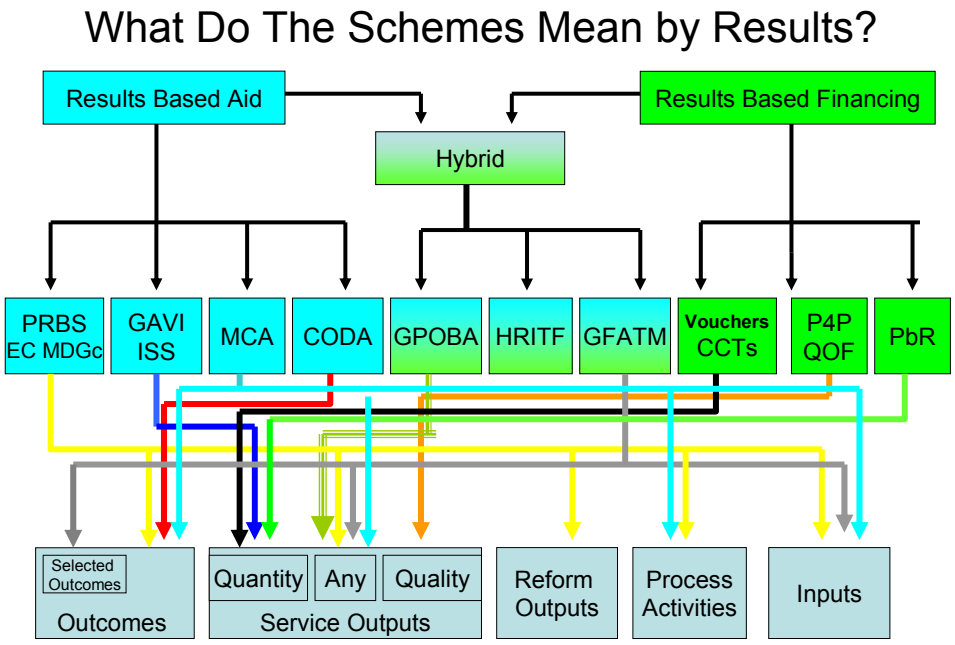


Chart 8 shows which types of results the various schemes are trying to achieve. It shows that results can be defined in:

- **Extremely narrow terms** e.g. quality – a particular dimension of an output – as is used under QOF in the UK and P4P in the US),
- **Somewhat narrow terms** (e.g. in terms of outputs (output based aid), or
- **Broad terms** e.g. Millennium Challenge Accounts and budget support under PRBS or MDG Contracts, where the results framework can be used to measure performance against a range of inputs, processes, outputs and outcomes). GFATM uses a range of country level indicators at all levels – but the *scope* is

ultimate goal – which is understandable when those goals are extremely broad and some would argue "woolly".

narrow as indicators are restricted to the three diseases and not across the sector as a whole.

- **Ex post or ex ante terms** e.g. for MCA the results are ex ante as results must be achieved *before* agreements are signed and funds committed. For budget support the core tranche is based on ex ante results – for the variable tranche it is ex post

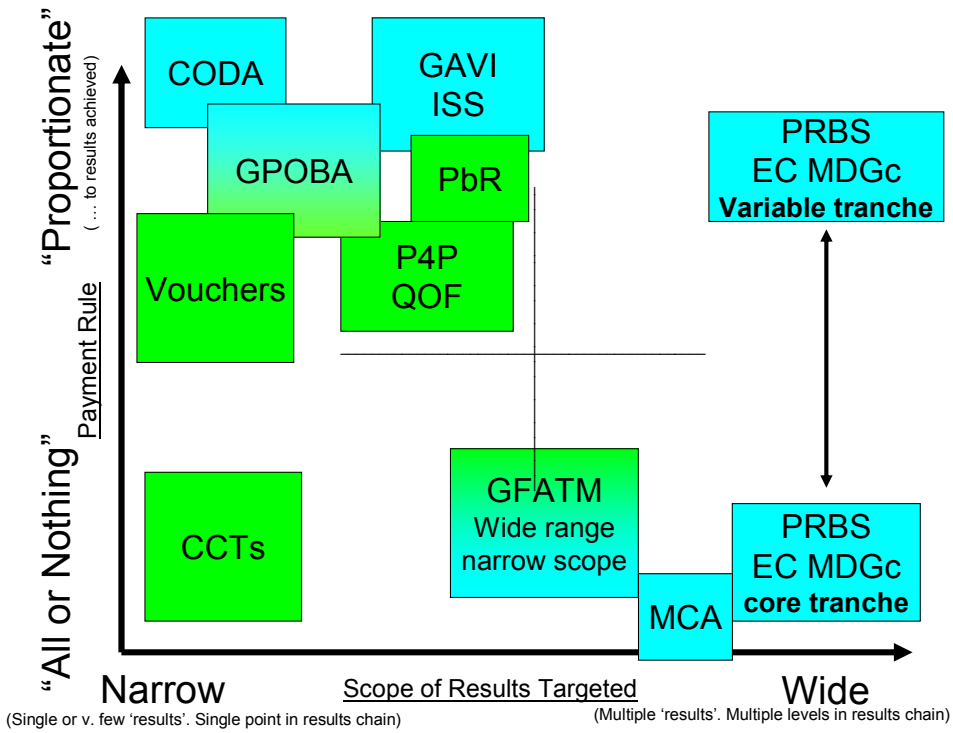
As the chart shows results tend to be set at the output level, though some focus only on specific components of outputs (e.g. quality) especially in developed countries where utilisation of services is already quite high and quality is the more pressing issue. Budget support and the Millennium Challenge Accounts are the only ones covering all levels including reforms.[14] It is also notable that none of the indicators in the results frameworks relate to aid effectiveness – although some approaches are inherently more consistent with the Paris Principles than others.

How the results are set. Again, this is less important in terms of the RBA/RBF distinction but does have relevance when one looks at consistency with the Paris Principles on Aid Effectiveness. In the case of PRBS and EC MDG contracts, the results framework is, in theory, mutually agreed between Government and donors. (If the donors have done their homework properly and Government is truly committed to achieving these results this does raise the question as to whether any performance based (variable) tranches are actually necessary – or whether it is just a public relations exercise on the part of the donor.). At the other extreme, GAVI ISS sets the results without any reference to country wishes. In between, countries challenging to access compact funds through the Millennium Challenge Account must deliver results in competition with one another against seventeen indicators that link to policies the government can influence within a two to three year horizon and that have a linkage (theoretical or empirical) to economic growth and poverty reduction,

How results are rewarded and the relationship between performance and rewards. In some cases the approach is “all or nothing” (the traditional way in which disbursements were made under the much-maligned structural adjustment instrument). In most cases the reward is intended to be *proportionate* to the *degree* of performance or progress towards the desired results. In some cases this is fairly easy to do as there is a direct relationship between the two. In other cases the link is not clear and any decisions about the level of funding are far more arbitrary with the focus being on achieving a fair outcome through dialogue.

[14] This raises an issue which is returned to below – if countries cannot scale up without reforming their institutional arrangements most of the mechanisms reviewed here are likely to be ineffective

Chart 9: Relationship between the types of results targeted and the rules for making payments



For example, if countries increase immunisation rates above the baseline rates they received funding in direct proportion to the extent to which they do so (i.e. by $20 per child). The link is clear and payment is automatic. Providers receive funding in proportion to the number of vouchers they succeed in attracting. Output based schemes pay a set amount for each unit of output delivered. CODA would pay out (ex-post) in proportion to results achieved (e.g. increases in primary school enrolment). CCTs, on the other hand, are only made after eligible families meet the agreed conditions so are of an "all or nothing" nature. In the case of the GFATM – although reprogramming is possible – the usual approach is to discontinue grants if performance is poor. Such decisions are based on a range of factors including contextual factors and the decision is administrative not automatic.

It can be helpful to break up some instruments. The fixed tranche of PRBS can be all or nothing and is based on ex ante partnership principle conditions being met (although it can also be *reduced* in value in response to particular breaches of underlying principles) whereas the variable tranche is intended to be proportionate to the (ex-post) results achieved against specific indicators, with relative weightings assigned to individual indicators (e.g. depending upon their importance for poverty reduction). The balance between the two varies – under EC MDG contracts the core element accounts for 70% of the total The whole programme, however, is underpinned by a shared agreement with government on fundamental issues (Partnership Principles for PRBS and Eligibility Criteria for MDC Contracts) and all support is conditional on adequate progress in these areas. In the case of the variable tranche, whilst there is an explicit link between rewards and specific outputs, the intention seems to be less to buy results and more to act as a signal to reinforce the importance of delivering results across the board and also to act as an incentive to strengthen data and reporting systems. Results are important – without them there would be no programme – but they are not driving the programme.

Progress under budget support is judged against a broad performance assessment framework. Unlike other schemes, poor progress does not automatically trigger a halt to payments (whether fixed or variable) – rather it triggers a process of policy dialogue which could affect disbursements but not until the following budget year at the earliest.

The 'odd one out' in some respects in terms of response to poor performance is the Millennium Challenge Account. This is not to say that results are not linked to performance – they are. Rather, it is the timing of the link that differs. The Millennium Challenge Account operates according to 'traditional' ex ante conditionality criteria. A certain level of overall achievement against seventeen indicators must be reached *before* a compact can be established. In a sense, this is similar to countries meeting PRBS Partnership Principles or EC Eligibility criteria – a certain minimum level of performance required for a (RBA) agreement to be established. However, there are two key differences. The first is uncertainty – a qualifying level of performance in one year may not be a qualifying performance in a following year if competing countries perform even better (and vice versa). It is all or nothing with a priori uncertainty. The second is that once a compact has been agreed, performance during the lifetime of the compact is not formally linked to performance (unless there is a fundamental reversal of achievement).

To summarise – for all the schemes results are important and continued disbursement will ultimately depend on performance (even if in some cases – MCA and fixed tranches of budget support – disbursements will only be suspended in exceptional circumstances). However, the ways in which payments are released and the response to lack of progress differs widely between programmes

The degree to which they are aligned with national systems. With the exception of some aspects of the Millennium Challenge Account, GFATM and, to a lesser degree, GAVI RBA schemes tend make significant use of national systems. PRBS and EC MDG Contracts budget support is fully aligned with national systems – with money flowing though the government budget and donors working with government planning and budgeting processes (such as MTEFs). The concept for Cash on Delivery Aid would also be aligned in the sense that there would be no direct impact on planning and budgeting processes and that funds would be used, accounted for, reported and audited entirely through existing government systems.

The Millennium Challenge Account requires partner countries to develop their own funding proposals consistent with their 'national development strategy'. In principle, Millennium Challenge Account money should be on-budget (i.e. appear in budget documents), but there is no guarantee of this in practice. Payments do not flow through existing government systems, although partner countries can designate an 'accountable entity' that will be responsible for implementing the Compact. The Millennium Challenge Corporation has yet to participate in harmonised diagnostic reviews of country systems. On a positive note, though, Millennium Challenge Account money is 'untied' – a major step forward for United States aid money.

A key principle of GFATM support is that its funding must be additional which, by definition, undermines national resource allocation processes. GFATM money is never on budget, rarely on plan and uses parallel processes for monitoring, accounting, auditing etc. Although GAVI money is usually channelled through national systems the funds are earmarked and there is no participation in national planning and budgeting processes.

RBF and hybrid RBA/RBF schemes tend not to use national systems, though this need not necessarily be the case. OBA schemes funded under GPOBA are normally contracted directly between the donor the (usually private sector) implementing agent,

thereby largely by-passing national systems. The UK and US models, on the other hand, clearly use national systems.

The degree to which the reward is financial or not.
Rewards are usually financial but, in the case of GFATM, this is not always the case. Good performers also have preferential access to fast-track light-touch follow on funding (through the Rolling Continuation Channel). Access to this channel provides additional funding but also reduces transaction costs. Whilst not directly associated with performance, PRBS and EC MDG Contracts are typically associated with a range of capacity building assistance. A Technical Assistance Management Agency (TAMA), for example, was established as part of the support package associated with DFID's National Health and Population Facility with the Federal Government of Pakistan (2003 to 2008). The aim was to source and manage national and international consultants to strengthen health and population welfare: policies and reform strategies; implementation capacity; management and organisation systems; service delivery; and knowledge systems and learning.

How the reward is calculated and the basis for calculation
With regard to PRBS and EC MDG Contracts, subject to donor resource constraints, the size of a budget support package and its planned disbursement over time are determined in large part by the parameters of the fiscal framework, including the policy for the fiscal deficit and the extent to which assumptions for donor grant receipts have been accommodated by commitments from other donor partners. The reward will equal the total size of the budget support package if all targets are met (including continued adherence to Partnership Principles/Eligibility Criteria), but may be lower if some targets are missed. In some cases, the size of a budget support package (or key components of it) may be influenced by the financing requirement for aspects of the poverty reduction strategy over the period of the support and by cost-sharing agreements with the government concerned. This was the case, for example, with one aspect DFID's health sector budget support in Pakistan to support the federal government in scaling up its Maternal and Newborn Child Health Programme, which commenced in 2007. The planned support was originally costed at around £90 million, accounting for around 28 percent of the estimated cost of scaling up services nationally in Pakistan.

With GPOBA-type schemes, the level of subsidy for a particular service (whether for a one-off connection to a service or an ongoing user fee subsidy) is estimated on a case-by-case basis. It may also be influenced by the procurement process (with evidence of contractually agreed subsidies in some cases being less than design estimates as a result of competition in procurement). With Cash on Delivery Aid, discussions in the literature suggest that payment could be in the region of $100 per additional student completing primary education and taking a 'nationally administered standardised competency test'.[15]

In the health sector, in some case payments are based on a fee for service. Under GAVI ISS, a country receives $20 per head for every child immunised above a self reported baseline. This was estimated to be sufficient to cover the actual cost of service delivery.[16] For GFATM the level of rewards reflects the size of the proposal which, in turn reflects the degree of ambition of the country in question. Under Payment by Results (PbR) the reimbursement is based on estimates of the national average unit cost of delivering the

[15] Based on the annual cost of putting a child through primary school in five developing countries, estimated at between $50 and $100, based on projections for average annual costs from 2003 to 2015 in 2000 US dollars. Millennium Development Project Needs Assessments: Country Case Studies of Bangladesh, Cambodia, Ghana, Tanzania and Uganda", UN Millennium Project, 2004.
[16] It seems apparent that the actual cost is much higher – especially at higher levels of coverage

individual services. With the Millennium Challenge Account, payment is based on the projected cost of schemes proposed by partner governments, subject to overall resource constraints.

The recipient of funds and use of funds

PRBS and EC MDG Contracts funds flow directly into the consolidated fund of the partner government and are completely unearmarked (although SBS conditions may stipulate some form of additionality in terms of level of spending for the sector concerned). Receipt of CODA funds would operate in a similar manner and to similar effect. Millennium Challenge Account funds are received by an 'accountable entity' designated by and acting on behalf of the Government. The funds are then utilised by the accountable entity on behalf of the government to implement the agreed activities. In some cases the funds end up in the hands of the beneficiaries (conditional cash transfers), in others with providers (vouchers; contracting of private sector organisation to delivery GPOBA-type projects), in others, the government (e.g. the Ministry of Health with respect to GAVI ISS). In some of these cases there are conditions on what the money must be spent on (GFATM – according to the budget set out in the agreement; Millennium Challenge Account – on the projects proposed to the Millennium Challenge Agency by the government). In most, however, (budget support, CODA, GPOBA, GAVI ISS, CCTs) the recipient can spend the money as it sees fit so long as the outputs are delivered as contractually specified (or, in the case of budget support, so long as there is 'satisfactory progress' in the implementation of the poverty reduction strategy, PFM strengthening and continued commitment to other underlying principles).

A final distinction is between **supply side and demand side approaches**. Most of the schemes support the supply side. Conditional cash transfers is the only scheme which focuses solely on the demand side. Voucher schemes combine demand and supply side elements (promoting competition between accredited providers to win earmarked in kind transfers).

Extent to which the scheme provides up front support to build capacity and help agents benefit from performance based payments.

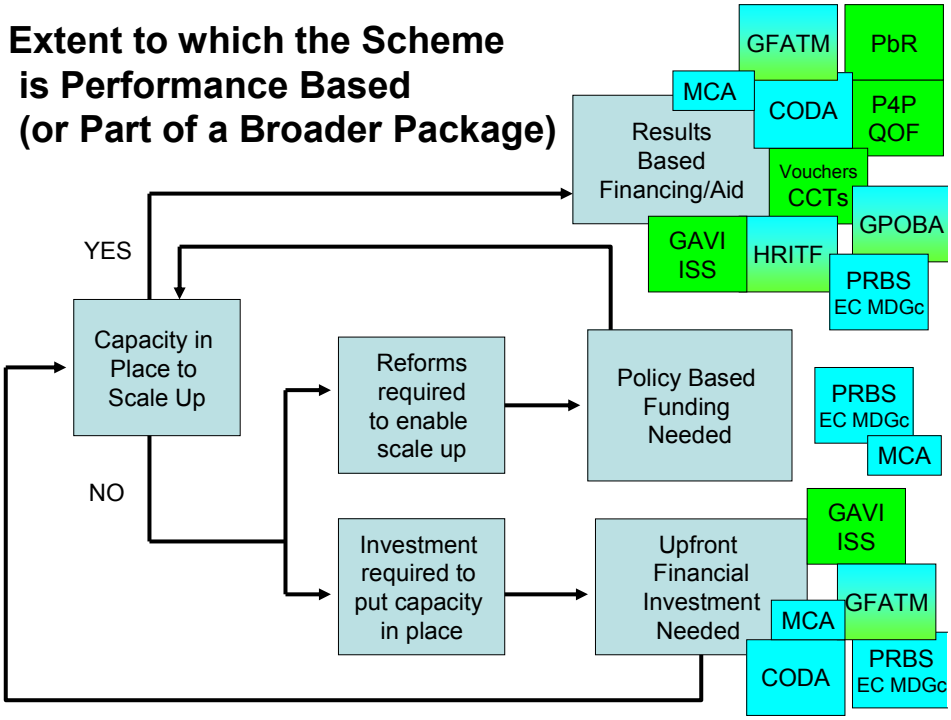
In the case of GAVI ISS initial funding is provided specially to help build the capacity to enable countries to benefit from the performance rewards. In other cases – e.g. PbR it is assumed that the capacity is there and the approach is "pure" RBF. In the case of PRBS and EC MDG Contracts *all* support is explicitly aimed at building long term capacity. Notwithstanding this, a typical budget support package will often incorporate a separate budget line for technical assistance specifically oriented to expanding capacities in particular areas. This would typically be accessible from the start of the budget support agreement and operate throughout much of the remainder of the agreement. This has been the case, for example, with PRBS support for education and health in Pakistan. Many budget support arrangements will also include resources for technical assistance for 'risk mitigation' purposes which are often oriented towards strengthening of weak PFM systems and anti-corruption.

Although there is a distinction between a fixed or core tranche and a variable tranche in budget support operations both are, in effect, performance related. The former focuses on high level and ex ante results (human rights, poverty reduction, etc.) and is aimed at ensuring a degree of predictability – the variable tranche tends to be more focused on a small range of more specific ex-post indicators or targets.[17] CODA plans to provide limited and earmarked funds to facilitate the development and enhancement of monitoring systems.

[17] Although future disbursements of fixed (and variable) tranches also depend on *ex-post* progress against the high level indicators.

It is not unusual for GPOBA schemes to include some up-front payments to address areas of capacity weakness that might otherwise constrain the effective delivery of outputs. Millennium Challenge Account *compacts* do not provide support that is oriented towards achievement against the seventeen indicators determining access to funds (unless compact-supported projects happen to coincide with such indicators, though this would only impact a future compact decision) but (non-compact) countries can apply for *threshold* payments that are oriented specifically to capacity enhancement against one or more of the seventeen qualifying indicators.

Chart 10



Reliance on competition and the focus on choice.
Some of the RBF schemes rely on competition to deliver results. Competitive procurement of OBA service providers (delivery agents) has been shown to impact on cost of schemes and delivery in some cases. In the Millennium Challenge Account, countries compete to gain access to funds. In PbR and voucher schemes providers compete on the basis of quality (price is fixed) to deliver services to beneficiaries.

In P4P in the US providers achieving better scores are rewarded with higher reimbursements than those who perform poorly. In other cases such as QOF rewards received by an individual provider are independent of those received by others. The concept of Cash on Delivery Aid would have little or no reliance on competition, with one proposal suggesting that an open contract be established for any eligible country to sign up to. Similarly, PRBS and EC MDG Contracts are not linked to competition in terms of determining eligibility.

A more detailed review of the differences between the schemes is provided at annex 6.

Key messages

- The distinctions within the RBA and RBF schemes add further complexity above and beyond the distinctions between the different schemes.
- These distinctions have major implications for how the approaches work as well as their consistency with the Paris Principles.
- Need to determine whether capacity is in place to support scale up – if not either specific measured will be needed or schemes may be ineffective.

5. Results measurement in RBA / RBF schemes

This section reviews the indicators and data sources used in major RBA and RBF schemes, identifies certain key measurement issues and makes some suggestions about how existing gaps may be filled, without undermining national statistical systems.

5.1 The indicators used

Different types of schemes tend to select different types of indicators. Narrowly-focused PBF schemes (such as GAVI ISS and the CODA Education proposal) usually have a single output or outcome indicator to reflect increasing service delivery or uptake (immunisation rates and the change in the number of primary school completers). In contrast, broader schemes (such as PRBS) often feature a long list of indicators at all points on the results change, from input and process indicators through to outcome. Booth and Lucas 2001 argued that there was often a 'missing middle' as the indicators the reflect the logical links between the two ends of the results chain were not there, but this aspect may be said to have improved since then. The GFATM falls somewhere between the two extremes: it 'typically focuses on five to 15 output indicators, as well as on impact and outcome indicators'

In some schemes common indicators are used (e.g. GAVI ISS, MCA); in others there is a process of negotiation to foster national appropriateness and ownership (e.g. PRBS, EC's MDG Contracts and the GFATM). Where common indicators are stipulated they are always clearly defined. This is generally also true where they have been negotiated: a good deal of time and effort in most developing countries in recent decades has been devoted to making sure targets associated with the indicators are SMART (Specific, Measurable, Achievable, Realistic and Timebound). (See also **chart 8)**

The suitability of selected indicators to act as a proxy for performance is a more complex issue. Various pitfalls that lie in wait for schemes that manage according to indicators have been identified as shown in Box 4. Annex 7 discusses them and then reviews the performance of the indicators selected by major RBA/RBF schemes and the way they use them.

Box 4: Key pitfalls of using indicators

Technical problems: Indicator not a sufficiently good proxy of the aspect of interest.

Tunnel vision: Concentration on areas included in the outcome-related performance indicator scheme to the exclusion of other important areas.

Suboptimization: The pursuit by managers of their own narrow objectives, at the expense of strategic co-ordination.

Myopia: Concentration on short term issues to the exclusion of long-term criteria, which may only show up in outcome-related performance indicators in many years' time.

Convergence: An emphasis on not being exposed as an outlier on any outcome-related performance indicator, rather than a desire to be outstanding.

Ossification: A disinclination to experiment with new and innovative methods.

Gaming: Altering behaviour so as to obtain strategic advantage.

Misrepresentation: Including 'creative' accounting and fraud.

Cherry-picking: (or 'cream-skimming') focusing on those near performance thresholds at the expense of those already above it or far below it.

Perverse incentives: where service delivery may ironically deteriorate in practice, rather than improve.

A mixed picture emerges. Mis-representation and tunnel vision emerge as the most common pitfalls encountered by PBA and PBF schemes. There is also evidence of perverse incentives and other gaming, cherry-picking, myopia and some technical shortcomings. In some longer running schemes, however, concerns have not always materialised as feared. The indicators for some RBA/RBF schemes appear to have been successful in avoiding these pitfalls.

In general terms, these issues are more problematic the fewer indicators are selected and the greater the stakes attached to achieving them. In particular it is hard for managers to alter their behaviour – or cheat – on a whole suite of indicators, and they are less likely to attempt to do so if the stakes are lower. Thus broader schemes such as PRBS, MDG Contracts and MCA are less susceptible to these pitfalls than focused schemes such as GAVI ISS and CODA. This is not to say broader schemes are immune: substantial funding may be strongly influenced by trends on a handful of key indicators, e.g. the incidence of poverty, especially where the payment of a performance tranche depends on them.

Perhaps better planning might have identified such adverse possibilities beforehand and allowed the design to be modified accordingly. Nevertheless, it is to be expected that RBA/RBF approaches will have unintended consequences, however carefully they are designed, especially when they are implemented alongside other reforms. The key issue is to identify the problems early and act accordingly. Strong monitoring and evaluation systems will play a key role in this.

5.2 The data sources used

A range of data sources is used in the different schemes. There are existing Government administrative systems and household censuses and surveys, and some of the broader schemes such as PRBSs and EC's MDG Contracts use them. Many narrower RBA/RBF schemes that may focus on a single indicator set up parallel administrative systems to try to generate reliable and timely data.

This section describes the various types of data sources that are (or may be) used in RBA/RBF schemes and assess aspects that affect their suitability for use in the schemes.

Government administrative sources

Government administrative data sources include Education and Health Management Information Systems. They are are generally collected by the ministry responsible for the sector. Professional statistical standards are generally much lower there than in national statistical offices. Statistical networks across ministries operate in few developing countries and they are seldom effective in influencing operations where they do exist. Statistical codes of practice are rare, and generally senior managers either do not understand or follow good statistical practice. Resources may not be sufficient to ensure robust data. In any case, the task of ensuring high data quality standards from health clinics and schools up through the system is very great.

There is plentiful evidence of data being misrepresented where resources are allocated, for example where capitation grants are calculated according to reported enrolment. In one particularly stark case:

> 'Kwara SUBEB [State Universal Education Board] conducted a detailed verification exercise into the implausible enrolment count in public primary schools provided by seven of the sixteen LGEAs in 2005-06. It concluded that the actual enrolment was under 100,000 rather than over 250,000 as reported. In the most striking case, one LGEA was assessed to have only one sixth of the total reported. [from ESSPIN 2009, based on Gannicott 2008, p 18].

Administrative data may also be flawed for innocent reasons. Schools are not always in a position to know various items on which they are asked to report, e.g. the age of children, whether new pupils are grade-repeaters or promotees, the resources allocated to the school and (even) their staffing complement. Teachers may lack a good understanding of what is requested and may not have the motivation to take sufficient care. There may be non-response, especially from the private sector.

Some audits are conducted to verify government administrative data. Simple headcount exercises to verify numbers of teachers and pupils are quite common in developing countries where there are concerns over misrepresentation. More detailed exercises are rare, though there is a promising system of sample audits of District Information System on Education. Independent Monitoring Institutes have conducted these audits in most States in India and the results are openly disseminated individually and synthesised (in Kaushal 2009 and Mehta 2008).

> 'The independent audit reports give a useful measure of the accuracy of the various data items and provide useful feedback on all aspects of the system, from the data capture format to the training given. There is scope for further improvements, including: reporting any bias – i.e. over-reporting or under-reporting – rather than net divergences; providing fuller meta-data to users on variables and States, and adapting the exercise to report on school coverage and response rates.' [See Ellison, 2009]

It is not known how costly these exercises are, but they probably constitute a small proportion of the total and may constitute good value in terms of the metadata generated and the future improvement of the exercise if the lessons are effectively applied. However, employing such (inevitably) small sample audits to affect reward payments on RBF schemes is liable to produce rough justice which may unfairly favour or penalise the beneficiary, depending on the units selected in the audit – and are thus likely to be contentious, as GAVI ISS has found.

Apart from suffering from data quality problems, Government administrative sources are often slow. For example, few developing country EMISs report within a year of the data being collected.

Administrative data are used in conjunction with population estimates**.** The denominator for key health and education indicators such as the immunisation rate used in GAVI ISS is the estimated population of the official relevant age group. Even where demographers use the best methods, because of flaws in the underlying source data such population estimates in developing countries cannot be precise, especially for infants and children. Even where censuses are regularly taken every decade (see below) population estimates will be based on a census taken between 2 and 12 years previously, as it takes about two years to release detailed data. The census data are likely to suffer from a significant undercount (e.g. an estimated 18 per cent in South Africa in 2001) and the reported age data is often very imprecise. Without usable vital registration data, occasional household surveys such as DHSs are used to update assumptions made about fertility and mortality in the projections. These aspects are subject to rapid change, as is migration within and

between countries. Population estimates, and hence immunisation and completion rates are subject to significant errors and occasional drastic revision. The effects are magnified as the estimated rates near 100 per cent. This longstanding problem does not appear to have an effective solution.

Parallel administrative sources

Government administrative systems are generally not considered suitable in narrower RBA/RBF schemes that directly reward according to single indicator values. The main reason is the susceptibility of government administrative data to misrepresentation, given the amounts of money involved on these schemes. The delays in data becoming available are another problematic aspect of government systems.

Establishing parallel systems gives a greater degree of control over aspects such as data quality and timeliness. In some cases, schemes devote considerable resources to M&E. For example, 'The GFATM recommends that countries invest 5 to 10 percent of their grant budgets in monitoring and evaluation.' The scheme's narrow focus also facilitates greater attention to detail: it is easier to control a single variable (e.g. immunisation rates) than hundreds of them that may feature on a typical MIS.

Even with the greater degree of control, the additional funds that may be applied to parallel administrative systems and the more intense focus, the quality of the data can often not be assumed. Schemes' data strategies often feature data quality audits. These audits play a vital role in verifying the results rewarded. The validity of GAVI ISS's audit system, however, was called into question by Lim et al 2008 who reported little correlation between the audit verdicts and the correlation triangulated household survey data, though the debate is complex and contentious.

Household surveys and censuses

Household sources are necessary for many indicators, especially at outcome level. In addition to population censuses, the following household sample surveys may be provide indicator values for broader schemes include: Demographic and Health Surveys (DHSs), Multiple Indicator Cluster Surveys (MICSs), Living Standard Measurement Surveys (LSMSs), Household Budget Surveys (HBSs), Integrated Household Surveys (IHSs), Expenditure Surveys (ESs) Income Surveys (ISs) and Priority Surveys (PSs). Taken together, they can yield valuable information. However, they suffer from various shortcomings when applied to RBA/RBF schemes: **periodicity, sampling errors and bias**. We discuss these in turn.

Ideally for such schemes, there would be annual data, as there usually is for administrative sources. This is far from being available from household surveys. The table in Annex 8 lists the surveys conducted in recent years for the top twenty recipients of bilateral UK aid in 2008/09 (which together accounted for 80 per cent of the total allocated to countries).[18]

We take as rough benchmarks that there should be:

- At least one general international household survey (i.e. MICS or DHS) every five years;
- One household survey that collects income or expenditure data every five years, and
- One population census every ten years.

[18] There is some conflicting evidence in the sources cited, but the overall picture is clear and will not be significantly affected by the small underlying discrepancies.

The table highlights the countries where there has been no such exercise conducted in the last five / ten years. The results are as follows:

- 5 of the 20 countries lack recent general household data;
- 12 of them lack recent household income / expenditure data;
- 7 lack recent population census data

The shortcomings tend to be larger in fragile and failing states, notably Afghanistan, Burma, Sudan and the Democratic Republic of Congo.

Second, all surveys are subject to sampling errors. These are quantified in DHSs and some other surveys. Sampling errors for some key variables at the national level for the 2008 DHS in Ghana are given in the table in Annex 9, as a fairly representative guide. For example, the 95 per cent confidence interval for the infant mortality rate runs from 62 to 84, which could easily lead to the reported direction of change from one survey being different from the true direction. DHS does not report sampling errors for maternal mortality rates, but UNSD 2003 describes them as 'very large' and as not being 'suitable for assessing trends over time or for making comparisons between countries.'

Third, survey estimates may be biased. For example, Carr-Hill 2009 estimates that accounting for the non-household population not covered by sample surveys (urban slum dwellers, street children, nomadic pastoralists and refugees) would add some 50 million or 65 per cent to the current estimate of 77 million for the out-of-school population. The generally consistent coverage of surveys from one round to the next means that the direction of trends is generally reliable, even if the level is not.

Surveys have not always been planned to minimise gaps between estimates becoming available and questions and definitions have not been harmonised. Various efforts are being made to improve the situation: internationally through the International Household Survey Network (IHSN) in response to Millennium Action Plan for Statistics (MAPS) and nationally through country-led National Strategies for the Development of Statistics led by the PARIS21 consortium. Also UNICEF says it will be prepared to provide assistance to countries at more frequent intervals – every three years instead of every five years. Nevertheless there are limits as to what is being achieved.

5.3 Summary of key measurement issues and proposals to tackle them

The TORs request this paper to 'identify the main measurement gaps in the different schemes and suggest how they could be filled, as far as possible using existing data sources and without undermining national statistics systems'.

The review of RBA/RBF schemes' measurement aspects highlight three main issues:

- How best to guard against misrepresentation and other pitfalls?
- How far investment in parallel administrative systems may be switched to supporting government administrative systems? And
- What to do about wider government statistical capacity and services?

These three issues are briefly described, before possible solutions are discussed.

How best to guard against misrepresentation and other pitfalls?
(Schemes where a major issue: GAVI ISS, CODA, Vouchers, CCTs)

Since the inception of the earliest PBA/PBF schemes, attempts have been made to design them to avoid the potential misrepresentation of data and the various other potential pitfalls identified. There have been refinements during implementation: schemes such as GAVI ISS and GFATM have generally been developed, implemented, evaluated by scores of professionals working in many countries for several years. In general, it appears these issues are addressed more or less satisfactorily in the established schemes. Nevertheless, the major issue today for several schemes remains whether improvements can be made to the data collection and audit procedures to ensure reliable data,

To the extent that such issues remain, it is unrealistic to expect a quick, general study such as this one to generate effective new solutions. The persistence of such issues in PBF schemes after many years in more propitious environments in the UK and USA would appear to confirm this view (although it should be noted the schemes' greater complexity present additional challenges). It remains to be seen if these issues can be tackled in a satisfactory manner in new schemes such as CODA, particularly in the education field as currently proposed,

How far investment in parallel administrative systems may be switched to supporting government administrative systems?
(Schemes where a major issue: CCTs, GFATM, GPOBA, Vouchers)

Various evaluations of RBA/RBF schemes have suggested they should support and work through government administrative systems rather than the parallel ones that they have used hitherto.

A number of observations may be made. In principle, it certainly sounds better to be investing in government systems that may be sustained in future, rather than parallel systems that currently undermine them. In practice, of course, it will be recalled the severe problems of data quality and timeliness that caused the schemes to set up standalone systems under project mode: there is often a long way to go before government systems may be suitable. The degree to which parallel systems really do undermine government systems may vary. It is hard to see how the collection of a single variable such as immunisation rates under GAVI or primary school completers sitting a terminal exam jeopardise existing comprehensive systems. (Were the schemes to replace other broader schemes such as PRBS and SWAps the picture would, however, be very different.) A scheme such as GFATM, in contrast, *is* big enough to displace attention, energy and resources that might otherwise go into broader Health Information Systems. Last, it is difficult to envisage schemes supporting government administrative systems beyond their area of direct interest which, as has been said, is often quite restricted. If so, it is unlikely they will be very successful in developing capacity to any great extent.

In summary, while the idea of mainstreaming parallel systems into government systems is worth investigating for various schemes, the problems may prove to be greater – and the gains smaller – than might appear to be the case.

What to do about wider government statistical capacity and services?
(Schemes where a major issue: PRBS, MDG Contracts)

PRBSs and MDG Contracts encourage and support developing countries to develop the capacity they need to produce the data required for M&E. There have also been a range of other collaborative efforts in the area. Nevertheless, there remain important gaps in the data available for many key partner countries, especially fragile states, as NAO noted in

its recent evaluation of PRBS. The gaps relate to the periodicity of household census and survey data and the quality and timeliness of administrative data, in particular. There is a need for further support if these broad PBA schemes are to be effectively monitored.[19]

The new Statistics for Results Facility may play an important role here. The Facility aims: 'to accelerate the strengthening of statistical systems in participating countries, so that there is a sustained improvement in the availability of and access to reliable statistics in order to better measure and manage for development results. It aims to help countries build the capacity to manage for development results, by linking national development plans and Poverty Reduction Strategies more closely to statistical efforts, and accelerating the implementation of country-owned statistical improvement plans.' (World Bank Website)

DFID is contributing £50m over 4 years. Afghanistan, DRC, Ghana Nigeria and Rwanda are the first pilot countries.

Key messages

- Despite careful design, many RBA/RBF schemes still face fraudulent reporting and their narrow focus can work against important broader aspects of progress. Narrow, high-stakes schemes are generally more strongly affected. The established schemes tackle these issues more or less satisfactorily, but some problems may be inherent and cannot be eradicated.
- Government administrative systems are often untimely and produce unreliable data. Narrower RBA/RBF schemes often set up parallel administrative systems which are generally more effective, though not perfect. Government household surveys are not conducted very frequently in developing countries and they suffer from sampling errors and biases, which limit their application to wider schemes. These problems with data sources tend to be worse in fragile states.
- While the idea of mainstreaming parallel systems into government systems is worth investigating for various schemes, the problems may prove to be greater – and the gains smaller – than might appear to be the case.
- DFID and other development partners such as the EC that operate broader schemes need to renew their efforts to encourage and support wider statistical capacity building. The new Statistics for Results Facility may play an important role here.

[19] Note: the need for data to monitor PRBSs and MDG Contracts may be seen to be secondary to Governments' own needs to support evidence-based policy making.

6. Evidence of impact and emerging lessons

Assessing whether RBA/RBF works or not

How would we decide whether RBA/RBF works or not? The first question one would have to ask is **"have the right results have been chosen?"** This is beyond the scope of this review but is a real issue nonetheless.[20] The next question would be "**does the scheme actually deliver results?"** – the question at the heart of this review.

The key issue here is **how to disentangle the effects caused by the specifics of a scheme – through its results focus – from the effects due simply to the additional funding attached** (and which could have been used to deliver through a different modality) and the effects due to entirely exogenous factors. Ideally, any interventions would be compared to a control in which the identical amount of funds[21] was channelled through an alternative mechanism.[22] Very few studies actually do this and for those that do[23] there is little hard evidence that any improved performance is down to the scheme itself.

The next question is that **"having identified the RBA/RBF specific benefits arising from achieving the desired results, do these benefits outweigh any possible costs associated with the scheme".** This reflects concern as to whether the costs of any unintended effects or related transaction costs have been taken properly into account.

Key findings

1. RBA and RBF generally do deliver results

There are many examples of schemes which have delivered results. The UK National Audit Office[24], drawing also on a recent major evaluation of budget support[25], reports that: budget support has been responsible for increased "pro-poor expenditures" 'in six out of nine countries' (with insufficient evidence in the other countries) has increased the quantity of service delivery in seven out of eight countries, usually in basic education or health. It also found that maintaining quality 'has proved challenging', with an expansion in basic services often being accompanied by a deterioration in quality (e.g. as governments seek to improve enrolment rates, pupil numbers may increase before the government has been able to recruit and train more teachers – as happened in Rwanda). Where macroeconomic stability existed beforehand, there is evidence that budget support has helped to reinforce it. See annex 10 for a summary of key findings from evaluation of budget support.

[20] Questions are increasingly being raised, for example, as to whether, too much aid funding for the health and population sector is being channelled to HIV/AIDS at the expense of other key areas such as maternal health (Shiffman 2009). Much of this funding is channelled to ARVs which the WHO CHOICE model suggests is one of the least cost effective use of aid funds. The implication is this is that however effective the mechanism donors may, quite simply, be funding the wrong things.

[21] This would include all funding associated with the RBA/RBF approach i.e. the direct funding but also indirect funding for example the additional transactions costs involved in supervising the scheme and monitoring and measuring progress.

[22] The integrity of any study would rely heavily on the identification of an appropriate counterpart.

[23] So far for just RBF schemes although such impact evaluations are currently underway in Millennium Challenge Account countries.

[24] 'Department for International Development, Providing budget support to developing countries', Report by the Comptroller and Auditor General | HC 6 Session 2007-2008 | 8 February 2008.

[25] 'Evaluation of General Budget Support' IDD and Associates, May 2006.

A recent World Bank literature review (Brenzel 2009) found that "RBF mechanisms appear to increase utilization of priority MCH services, and that CCTs have even been shown to have positive effects on health child outcomes" It found that performance based contracting was *associated*[26] with increases in service utilisation and better access for the poor (in Cambodia and Haiti with "substantial increases in the use of family planning services, institutional births, assisted deliveries, and DTP3 coverage" in the fee for service based system in Rwanda) but not in Uganda.[27] It found conditional cash transfers had "a marked impact on utilization of essential preventive child health services "and had good equity characteristics.[28] Vouchers also "appear to have had a positive impact on utilization of heath services".

The recent Global Fund evaluation suggested that "increased funding (for HIV/AIDS) is resulting in better availability and utilization of services which ultimately will have an impact on disease burden" but that there was "no clear trend in levels of access to TB services. The GAVI ISS evaluation found that ISS had "a significant positive impact on DTP3 coverage rates from 2001-2005. Some RBF approaches were found to improve the efficiency of services "because they only entitle recipients to a defined set of services" and reduce incentives for over prescription Bhatia & Gorter, 2007). At the same time it found that the impact on quality of care was unclear and that there were a number of unintended effects. US P4P approaches have found at best a modest improvement in quality compared with non participating hospitals – with many studies showing little or no impact. Possible reasons for the failure to see more impact might include delayed response to the incentives[29] or the fact that reward money was not necessarily immediately reinvested in services.[30]

A range of notes produced by GPOBA as part of its dissemination strategy on OBA schemes showed that the schemes generally performed well against the results anticipated in their design.[31] Notwithstanding the application of OBA-type schemes since the 1990s, however, the database of evidence describing their impact compared to alternative methods of achieving the same objectives is sparse. For instance, a 2009 OBA review[32] showed that a total of 85% of OBA projects achieved or over achieved desired results within or below budget, compared to 49% of traditional projects. Whilst this appears to indicate relative success, the 'desired results' of each as referred to in the report are not directly comparable (in the sense of one being a control group for the other). Without further information, it would be impossible to say, for example, whether

[26] The use of the terms "associated with" rather than "responsible for" is crucial.
[27] The authors speculated this might be related to the small size of the bonus, complexity of the contracts, or short duration of the incentive period.
[28] In Nicaragua, the CCT increased immunization coverage for children traditionally hardest to reach, who live farthest from health facilities, and whose mothers had not completed primary school.
[29] Early assessments of the impact of PbR on provider activity in England found no strong evidence of this (Audit Commission 2005, Farrar et al 2005). However, some researchers have concluded that this may be because organisations were still familiarising themselves with the system (Farrar et al 2005).
[30] The GAVI ISS evaluation found that countries "do not necessarily spend all (or even much) of the reward funding in the year it is received, nor in the following year. In all four countries, the reward funding received was quite substantial compared to prior years' ISS funding and to total immunization program expenditures. In Tanzania and Zambia, partners strategically decided to save some of the ISS money for later years or in case of future "emergencies" – this approach was not specific to reward funding, but was true for management of ISS funding in general in many countries".
[31] E.g. OBA Approaches notes: number 17 (Connecting poor households to natural gas service in Columbia); 18 (Expanding telecommunications services to rural areas in Mongolia); 25 (Extending Water Services to the Poor in Urban Areas in Morocco); 26 (Expanding Water Supply Service in Rural Areas in Morocco); 28 (Improved Access to Water Services for Poor Households in the Philippines); 29 (Facility for the Water and Sanitation Sector in Honduras); 30 (Piped Water Infrastructure in Kenya
[32] 'A Review of the Use of Output-Based Aid Approaches'. International Development Association Global Partnership on Output-Based Aid (GPOBA) Finance, Economics and Urban Department, Nov 2009.

the data simply reflect a more realistic approach to target setting than may otherwise be the case.[33]

Similarly, information collated by the Millennium Challenge Corporation demonstrates that results are achieved through their projects in compact countries (e.g. in terms of kilometres of road construction completed, hectors of land under agricultural production), but there is no evidence yet as to whether these results are different to those that would have been achieved by devoting the same overall resources to an alternative aid modality, although a range of impact evaluations currently under way promises to address this through control group analysis.

Unintended effects do appear to be widespread. The GFATM evaluation, for example. found that "basing the GFATM's PBF (performance based funding) system largely on numeric output targets created unintended negative consequences, especially in terms of the quality of service provision. Implementers in more than half the SA2 countries reported that, on at least one occasion, they had sacrificed quality of implementation in order to achieve a quantitative numerical PBF output target".

2. Attribution of results to RBA/RBF is generally not possible due to the lack of well designed evaluations

The evidence base supporting results based approaches – especially in low and middle income countries – is extremely weak[34] (Oxman).[35][36] Study designs rarely allow systems impact to be assessed – controls are rarely included[37] Eldridge and Palmer found "a lack of clear evidence on the effects of any type of PBP (Performance Based Payment) in any low income country health setting. This was largely due to the absence of controls in most studies; the only study to include control sites found that they outperformed those with performance based payments (PBP) (Lundberg 2007). Despite this, most of the papers reviewed provided a favourable assessment of PBP".[38]

There have been no control studies of pure RBA schemes, although there has been a recent major evaluation of budget support which, on balance, shows that results have been positive and unintended consequences (particularly in terms of aid effectiveness) have been small. Constructing a counterfactual for the relative impact of aid resources provided through general budget support presents a much greater challenge than would be the case, say, assessing the impact of a more precise intervention from a health RBF scheme or a utilities hybrid (RBA/RBF) scheme. Even so there is still some evidence

[33] The contractual implications may, in fact, provide an incentive to an implementer to ensure realism that may not exist in an environment where the accountability for delivery of project outputs may not be so strong.
[34] There is a lack of robust evidence regarding effectiveness of various mechanisms and their impact on health status in low-income countries: Brenzel 2009
[35] Oxman, Andy. An overview of research on the effects of result based financing. Report from Norwegian Knowledge Centre for the Health Services nr 16 –2008
[36] According to the *First Evaluation Report to Congress* on US P4P initiatives in 2006 ": the lack of up-front payment for the investment in new systems of case management "have made for an uncertain future with respect for any payments under the demonstration."
[37] In the US, for example, early excitement about improvements from P4P experiments were tempered by the realisation that improvement was taking place across the board not just for those providers benefiting from P4P.
[38] Eldridge and Palmer emphasise the "lack of controls and the interference of confounding factors. It is not possible to isolate the effects of PBP in Cambodia and Haiti because no data were collected from control sites In Cambodia, the evaluation of the PBP was not isolated from other variables such as (1) the total increased salaries for practitioners, (2) the increased financial support from a loan from the Asian Development Bank to the Ministry of Health for technical advice and management, and (3) the effects likely to be created by the arrival of an international NGO in a district, such as the increased presence of doctors In the only study in which adequate control were in place "facilities in the control group outperformed those receiving PBP on several indicators (Lundberg 2007)".

which suggests that the impact of budget support on health and education service delivery is generally positive as noted above. At the same time the complexity of the transmission mechanism (and the time frame required for impact to be measurable, not to mention the conceptual issues associated with the choice of poverty indicator) means that there is still little solid evidence of results in terms of changes in poverty levels that can be directly attributable to budget support.

The Millennium Challenge Corporation is currently undertaking research on Millennium Challenge Account interventions in compact countries which is meant to include control groups where possible, although none of these studies have been completed to date (though there is plenty of information on what has been delivered to date). The concept of Cash on Delivery Aid has yet to be piloted, so there are no results to present. The design of the Millennium Challenge Account interventions intends to overcome this through analysis of impact in the context of results for a control group where this is possible, results of such analyses remain forthcoming.

Several studies refer to the lack of evidence on the "relative cost effectiveness compared to other approaches to increasing uptake by the poor", although, to be fair, this finding applies widely across the aid environment and is not exclusive to RBA/RBF.

Thus, whilst much of the evidence that is available does appear to show improvements in outputs (and sometimes in outcomes), **the evidence is generally unable to demonstrate that the results achieved can actually be attributed largely to the results based arrangement put in place or whether achievements would not have equally successful and/or costly with an alternative means of achieving the same set of results.**

3. The presence of complementary reforms makes attribution even more challenging

Most of the evidence that is available tends to suggest that complementary reforms are often required to put in place the necessary building blocks for RBF and RBA to be effective. For instance, there is little point in addressing demand side factors if the supply side is not able to respond.[39] Rwanda, for example, has embarked on a series of wide ranging reforms aimed at both supply and demand side including decentralisation and autonomy of health units as well as payment reforms. UK efforts to strengthen GP services involve 5 key interventions with 13 objectives.[40,41]

The provision of general budget support to Rwanda enabled school fees to be abolished in 2004 resulting in a corresponding expansion in primary school enrolment from 73% (in

[39] In the case of the maternal voucher scheme, in Bangladesh there have been reports of mothers waiting long periods of time to be reimbursed for transport costs.

[40] The New GP (GMS) contract involved the provision of a range of incentives and payments (a needs based global sum, incentives to provided additional services and cover out of hours, quality incentives through QOF and staffing incentives to promote new skill mixes. This package was intended to deliver better productivity, more needs based allocations, greater choice, better infrastructure, more appropriate skills mixes, better quality of care and better clinical governance, lower administration costs, reduced demand for secondary care and capacity of primary care physician to expand services provided, increase the number of GPs, develop a better career structure and promote more flexible employment patterns.

[41] It is interesting to contrast the UK experience under QOF with that of PbR. In the case of the former, it has been argued that earlier investments put in place the framework in which RBF could succeed. Thus, on the face of it, the marginal investment associated with QOF seems to have had a major impact (the provider response has been much more rapid than the Department of Health expected). Experience also highlights the importance of financial management systems.

2000) to 96% in 2007. This demand measure was put in place before supply side arrangements (e.g. recruitment and training of additional teachers) was possible. Enrolment expanded at the expense of quality. In the UK the NAO considered that PbR had (alongside a range of other measures) "at most, *contributed* to a range of positive trends rather than *driven* them".[42]

These experiences suggest that if schemes are established to expand service delivery they should go hand in hand with complementary measures to ensure results are achieved. It might also suggest that whilst it might be easier to attribute results from stand alone RBA/RBF schemes – such schemes are less likely to deliver results in the first place.

4. Transaction costs can be extremely high (often too high).

(And there is a risk they will outweigh the benefits of any additional results delivered).

Transaction costs are rarely well defined. The literature provides no clear consensus; the DAC Glossary of Key Terms in Evaluation and Results Based Management provides no guidance on the issue. Some of the literature considers all institutional costs – both internal and external to the principal and agent – as transaction costs whilst other approaches tend to exclude internal costs.[43] Given this lack of clarity and the fact that costs, however defined, are not tracked it is not possible to answer key questions such as:

- How high are transactions costs?
- Who bears the costs – governments or donors?
- Which costs are "good" (i.e. necessary and add value) and which are "bad"? (unnecessary and add no value)
- How transaction costs for one instrument e.g. RBA/RBF might compare to that of any other e.g. technical assistance and
- Whether there are economies of scale according to the size of programme

DFID might wish to consider commissioning some work on such issues – perhaps through one of the existing Research Work Programmes.

It is generally *assumed* that RBA/RBF scheme related transactions costs will decline over time and be lower than alternative approaches. There is little evidence that the expected decline takes place – certainly in the short term. Costs can be high – often unnecessarily so (see box 5) – yet the evidence suggests they may still not achieve the level of oversight required. Little can be said in terms of comparisons with other instruments.

[42] Attribution is also made more difficult when one considers that care was already improving rapidly in the years leading up to the new contract. The percentage of patients with controlled blood pressure rose from 47% to 72% between 1998 and 2003 and the percentage of patients with cholesterol within recommended levels increased from 18 to 61 % in the same period. The reasons for these improvements go much further back and include audit which introduced as a compulsory part of the 1990 GP contract.

[43] Cheung includes any costs that arise due to the existence of institutions; others such as Demetz (2003) and Werin et al 1992 exclude internal costs and only consider costs associated with the market mechanism/contracting arrangement.

Box 5: Transactions costs can be huge. In Rwanda, Olsen reports that the "validation of data is comprehensive and time consuming, illustrated by the example of a team consisting of five district health professionals using an average of five hours each per month just to validate data in one health centre".[44] The failure of schemes to align with Government processes, should be relatively straightforward to rectify in some cases but not in others.[45] Efforts to strengthen monitoring systems have often created major inefficiencies and unnecessary costs. For example, the GFATM evaluation found that "the increased demand for information by most development agencies, combined with a lack of strong country-led plans, has contributed to a perception in countries of being "overrun" by uncoordinated surveys, incompatible and poorly supported information technology solutions, different clinic-based reporting styles related to HIV/AIDS, and multiple demands for donor reports" and that "fragmentation along disease lines has created problems of overlap and duplication at country-level. Surveys targeting the three diseases have been implemented in a poorly coordinated way by different donors. The rush to strengthen clinical-based reporting systems to address the long-term management of people on antiretroviral therapies has resulted in the introduction of multiple information technology solutions including electronic medical records, and a range of non compatible hardware and software".

However, even allowing for the large burden it is far from clear that the systems put in place to track results are fit for purpose. The GFATM evaluation found, for example, that "while the system generates extensive data, it often fails to provide the key elements of information required to inform judgments on effectiveness". Furthermore whilst it commended the Fund on its efforts to improve the systems, this had only had the "unintended consequence of making the system more confusing at the level of implementation". Although popular amongst development partners the Fund model was seen by many countries as" burdensome, rigid, and fixed exclusively on short-term outputs rather than on longer-term outcomes, results, and capacity building". Funding some or all of the transactions costs and even better providing Governments with the resources to cover them seems to be popular and work well.[46]

There is a risk that schemes such as Cash on Delivery Aid, the Millennium Challenge Account, GPOBA and the expanding array of RBF and hybrid RBA/RBF schemes become just one more layer of instruments that require dialogue, negotiation, reporting and verification in addition to an existing array of projects, SWAPs and budget support. More specifically, although RBF and hybrid RBA/RBF schemes typically require government approval (as is the case with OBA schemes operating under World Bank procedures, including those funded through GPOBA), this is no different to existing requirements under partner government legislation and procedure (in Pakistan for example) for project approval (off-budget or otherwise)

Different aspects of transaction costs can also move in opposite directions. One of the key objectives of budget support design is, in fact, to lower transaction costs, particularly

[44] http://siteresources.worldbank.org/INTHSD/Resources/topics/415176-1217510876610/5259033-1226502367411/RWANDAHEALTHSYSTEMPBFANDMDGS.ppt

[45] In Vietnam, for example, a GFATM evaluation stated that "all national health/donor reporting is aligned with Ministry of Health quarterly and biannual reporting, except for GFATM quarterly financial and activity reporting on a TB grant. The latter's financial report is off cycle by just one month, resulting in the need to recompile all the quarterly financials rather than use existing data".

[46] There has been a recent tendency for global health partnerships to pass on transaction costs to country level partners – at the same time satisfying their Boards that they remain lean and mean. One interesting feature – and which is an example of good practice – is the willingness of the GAVI HSS window and the World Bank's HRITF to internalise such costs by providing project preparation grants or seed funding. Such flexible funding, though not without problems, has been popular with countries.

for the partner country in managing its relationship with donors. In this, the evidence shows that budget support has been a success – the recent major evaluation of budget support concluded that governments' transaction costs at implementation were significantly reduced by virtue of being able to follow standard government procedures rather than a multiplicity of donor ones. The evaluation also concluded, however, that up-front transaction costs were not perceived to have fallen as much as 'some had expected', whilst the UK NAO's own survey of DFID country teams suggested that more staff would be required in the short term to manage budget support programmes effectively. Neither report concludes which of these changes in transaction cost is most important, although it would be reasonable to expect the fall in partner country transaction costs to be more substantial. (It does suggest, however, that these additional transaction costs should be explicitly catered for by budget support providers, notwithstanding pressures to reduce administrative cost whilst expanding aid budgets).

Targeting is an important aspect of design for many RBF and hybrid (RBA/RBF) schemes in particular with implications for both equity and cost. The literature on targeting demonstrates a straightforward trade-off between precision and cost. Many OBA schemes, for example, use geographic targeting which minimises transaction cost but is open to leakage. If the target group is concentrated in a particular area (e.g. the poor in an urban slum), the transaction costs of targeting will probably be low. As efforts are made to reach other members of the target group, transaction costs inevitably rise (although this is a feature of all aid modalities targeting the poor). Having noted this, there is, in fact, little or no information in GPOBA annual reports, reviews or other OBA documentation on overhead or transaction costs (whether associated with targeting or otherwise). Whilst the conceptual approach suggests that such costs should be relatively low, this has yet to be demonstrated.

5. There are questions as to whether any results being achieved by the schemes will be sustained

This reflects a number of concerns:

i) RBA/RBF may simply not deliver in the long term There is a risk that **agents may adapt** their approaches in over time in light of experience and find ways of making economic rents without accepting a reasonable share of the risk. In many cases the approaches being adopted are new and are not yet fully mature or have not yet been fully rolled out so this has not been fully tested. Oxman, for example, concludes that "financial incentives seem to be effective in the short term for simple[47] and clearly determined objectives but argued that it is less clear whether financial incentives can provide longer term changes". A general impression in the UK is that providers are continually able to keep one step ahead of those commissioning services.

ii) RBA/RBF schemes may deliver but the wider costs may exceed the benefits There may be trade offs between a results focus and the "other" Paris Principles. If the Paris Principles are seen as a key to sustainable long term development (which itself may be open to some doubt) there is a risk that focusing on one aspect of these principles to the neglect of the others may not serve this purpose well. There is little or no evidence in the literature on the extent to which many of these RBF and hybrid RBA/RBF schemes (OBA for example) form part of a harmonised sector-wide approach. There will inevitably be some temptation to by-pass much of the bureaucracy associated with this If

[47] Doolan 2004 suggests that RBF approaches in the US have been more effective where the intervention requires less patient cooperation.

not by-passed, these schemes then become one extra layer to manage within a SWAp or similar mechanism.

By way of contrast, budget support has been explicitly designed to be consistent with Paris Principles and the Accra agenda and, evaluations demonstrate that it is, in fact, broadly consistent with these. Not only does the evidence demonstrate that it is consistent with government systems and actively supportive of government policies, there is evidence to demonstrate that it has strengthened the capacity of government systems (e.g. through PFM reform). The Millennium Challenge Corporation has made efforts – with mixed success – to ensure its support is consistent with Paris Principles. Compact countries propose the schemes to be funded according to their own requirements which is important for alignment. Governments specify their own 'accountable entity' that is responsible for managing funds and are encouraged to present them in budget documents, but the funds must be separately accounted for and auditable and do not, therefore, operate through government systems.

In practice, many RBF and hybrid RBA/RBF schemes include features of the traditional project approach (aside from by-passing government systems). Experienced staff, for example, will inevitably be tempted to migrate from relatively poorly paid government positions (although this has yet to be addressed by research associated with such schemes), thereby adding to existing constraints to effective reform of staffing, management and pay structures.[48] These harmonisation issues can be partly addressed by attempting to ensure that procedures for design of schemes take them appropriately into account. To some extent, however, it will probably be necessary to accept that many RBF and hybrid RBA/RBF schemes will not be well harmonised. The issue, as noted earlier, then becomes one of whether the results are worth the costs (including loss of harmonisation).

Contracting out service delivery represents a good example of the potential conflicts. By definition, RBF and hybrid RBA/RBF approaches that implement through non-government actors (which most do) do not use government systems for procurement, accounting, (some aspects of) reporting and audit. To the extent that this represents, in effect, the contracting out of a service by the government, so far so good – so long as capacity and skills are being created within government to engage in such contracting-out (there is not much evidence for this). In Cambodia, for example, the shift from contracting out approaches to an internal contracting model has been hampered by the fact that the provincial health departments now responsible for managing contracts had little or no role in the earlier model. The contracting out model is now widely used in fragile states and countries emerging from conflict and this issue is likely to reoccur.

iii) RBA/RBF may take the pressure off and reduce the need for key reforms

Finally, though not necessarily subject to testing, there is a risk that some RBA/RBF schemes may act as a short term "sticking plaster", reducing the need for more fundamental or systematic reforms which would support long term sustainable development (e.g. does salary supplementation through PBF reduce the demand for civil service pay reform).

6. The equity picture is mixed

Equity can mean many things and can be looked at from a number of perspectives. Equity of what – inputs? outputs? outcomes? Are we talking about equity *between*

[48] The World Bank OBA Manual advises scheme designers to pay bonuses to health sector staff to combat resistance to schemes.

countries or *within* countries? At what stage is inequity generated – approval stage? implementation?

To a degree equity is governed by the range of services to which the scheme is targeted.[49] This, in turn, may be governed by the institution actually implementing the scheme and their particular mandate. In practice, institutions with a relatively narrow focus – such as GFATM and GAVI – have been at the forefront of RBA/RBF.As a result, equity may be driven less by the RBA/RBF approach itself than the institutions that implement them.

The GAVI ISS evaluation found that whilst there was little evidence of growing inequity *within* countries Low Income Under Stress (LICUS) countries "tended to apply for GAVI ISS funding much later than other countries…and were also less likely to receive reward funding". The team hypothesised that "a fragile government responds differently to an incentive-based funding mechanism than a well-functioning established government" and argued that GAVI should investigate alternative approaches for working with fragile states, including more involvement and up-front assistance. It warned that "if ISS funding remains a primarily performance-based funding mechanism, funding by definition will go primarily to better performers, which are unlikely to be countries under stress". The Millennium Challenge Account specifically channels resources towards countries which have *already* demonstrated good performance thus explicitly excluding poor performers which are typically those with the greatest needs. GFATM, by contrast, has found that performance-based funding does not disadvantage poorer countries (with grant performance in low income countries just as good as those in high income countries (according to the share of grant rating B2 or C).[50]

Targeting (of poor groups) is possible but involves costs and significant implementation difficulties.[51] Though there are risks that poor countries and the poor within countries will be disadvantaged though this is not inevitable and fears that it would occur have often proven to be largely unfounded. Reviews of QOF in the UK, for example, reported good results in deprived areas.[52] The GAVI ISS evaluation found "no correlation between rewards and geographic equity or stability of coverage". The GFATM evaluation found "no evidence of widening or narrowing gaps in coverage between disadvantaged groups and those who are better off" but it did identify major gaps in the monitoring of gender, sexual minorities, urban-rural, wealth, education, and other types of equity as part of grant performance.

[49] The equity impact will often be driven by the nature of the service being targeted. Where coverage of higher income groups is already almost universal (e.g. immunisation), increased coverage will automatically benefit lower income groups. For other services such as ART where coverage is lower and restricted to urban settings, this is less likely to be the case.

[50] The Fund put this down to the fact that "performance-based funding approach does not disadvantage countries with the lowest income levels or weaker health and community systems (performance is measured in relative terms against the unique targets established by the countries".

[51] Targeting is possible but it can be time consuming and expensive. In terms of demand side financing initiatives, the approach was usually to target specific services or target by geographic area, often providing universal coverage in poor areas and means tests in other settings e.g. Universal Coverage in Nepal and less developed states in India, with poor and scheduled castes and tribes targeted in other Indian states. Schemes usually do succeed in raising utilisation in target groups. For example Borghi et al. 2005 found positive effects among the target group regarding utilization of vouchers for key preventive health services. However, such schemes also face substantial costs and implementation difficulties, including administrative complexity and costs, fraud and leakage of subsidies to better off groups, costs of selecting target groups.

[52] Several research groups have found that QOF scores are lower in deprived areas. But the interpretations of this finding are very different. One conclusion is that GPs in deprived areas achieved high scores without recourse to high rates of exception reporting, and the differences in scores between affluent and deprived areas are small and of relatively little clinical significance. This is a considerable achievement for practices in deprived areas".

As RBF and hybrid RBA schemes must specify the outputs that are to be delivered, beneficiaries can typically be more readily identified than in traditional inputs-based schemes. The November 2009 review of OBA schemes notes that an explicit targeting mechanism was used in 97 out of 127 projects (with 23 of the remainder being road infrastructure projects for which targeting was less relevant). More than two thirds employed geographic targeting (e.g. water supply in slum areas of Kampala). 23 per cent used self selection – providing higher subsidies for more basic services (e.g. subsidising solar home systems in rural Bolivia and external yard taps in Mozambique). 10 per cent used means testing – mainly in middle income countries (e.g. access to gas and heat supply in Armenia). Community-based targeting was also used, but only in a few cases (e.g. water access with small-scale providers in Cambodia). The evidence with regard to OBA, therefore, demonstrates that targeting is, in practice, an integral component of the vast majority of the schemes. Moreover, given the nature of the targeting and the type of output delivered, it is likely that many if not most OBA outputs (e.g. improved household access to utilities) will be accessed relatively equally by men, women and children. Beyond this, what the evidence does not tell us the extent to which targeting has been successful (e.g. in terms of the proportion of beneficiaries that met the targeting criteria – usually pro-poor) and the additional cost associated with targeting. Given the relatively small-scale nature of the projects, however, targeting for most of them (geographic, self-selection) is likely to have been both relatively successful and cost-effective. This will change on both counts, though, if and when OBA projects become larger and more ambitious.

By definition, PRBS and EC MDG Contracts are expected to result in improved access to public services by poor people (a poverty reduction strategy being a prerequisite and continued support dependant upon satisfactory progress in poverty reduction). Evidence from budget support evaluations appears to confirm this (e.g. increased pro-poor expenditures in six out of nine countries surveyed). Evaluation of budget support has not focussed explicitly on its impact on gender equality. The approach taken in the evaluation, however, was to assess the extent to which poverty reduction strategies and budget support arrangements addressed gender (thereby presenting evidence for likely impact).

All countries address gender issues in their poverty reduction strategies to some degree, though some (e.g. Uganda, where there has been efforts to mainstream gender issues with budget support indicators) reflected gender in their budget support indicators more than others (e.g. Rwanda where gender was not explicitly addressed). The Millennium Challenge Corporation operates according to a gender policy that includes a commitment to gender equality, though data has still to be presented to demonstrate the extent to which this is being achieved through project in practice.

A number of features may lend themselves to more equitable results. These might include:

- The use of locally identified and relevant targets – which may be lower, but still challenging, in low income countries – rather than global targets or standards.
- The use of stronger incentives for poorer countries.[53]
- Up front capacity building efforts to help develop capacity to help weaker countries take advantage of RBA/RBF schemes.

[53] GAVI ISS unintentionally (?) provides stronger incentives for countries with the poorest coverage rates as the unit cost of providing immunisation tends to be much lower – similar to the reward level – in poorer countries. In higher income countries the unit cost of delivering immunisation is likely to far exceed the reward on offer.

- Different (simpler) approval processes and technical support may also be needed to ensure capacity constrained countries can benefit from RBF and hybrid RBA/RBF schemes in the first place.

7. There are questions about applicability of many RBF schemes to low income countries

Many of the approaches in use in middle and high income countries may not be applicable or appropriate in low income settings. Brenzel concludes that CCTs "are complex and present several management challenges". In view of the complexity of the institutional arrangements and the potential high cost of the transfers, there is some question as to whether approaches such as CCTs can be scaled-up and sustained in low-income countries. The issue is not just about funding. Despite having much stronger systems RBF was still seen as inappropriate for some services in middle income/developed countries.[54] The implication for this, again, is that the appropriateness of schemes in particular settings should be determined on a case by case basis

8. Politics matters

RBA/RBF schemes are far more likely to succeed when there is strong leadership and political commitment at the national level (Brenzel 2009 – see also **Box 6** below). Key decision makers may support RBA/RBF schemes for a variety of reasons. They may support them because they actually want to deliver the results agreed in the schemes. Where this is the case one would expect the prospects for the schemes to be good. In some cases they may be supporting the schemes only because they help achieve other objectives which also happen to deliver results. For example, schemes may be seen as a way of supplementing salaries and keeping health workers or teachers happy rather than through any great desire to achieve results. Should we worry too much where this is the case? Maybe we should? What if countries find other ways of supplementing salaries or achieve their objectives *before* results are achieved.

The key points are that politics matters, that political analysis of what drives politicians is often rather weak and that we should not take it for granted that achieving results is the main driving force. Where this is not the case there are questions as to how sustainable the schemes will be; where there is no political support it is doubtful that schemes could function effectively at all. This would emphasise the importance of a sound political analysis as part of the design stage.

9. Good design is essential … and may help to reduce unintended effects

As with many reform initiatives "the devil is in the detail". The design of a typical partnership general budget support intervention, for example, is based on many years of experience, research and policy thinking which is reflected in a range of detailed policy documentation that explains clearly, inter alia, what the purpose of budget support is, the principles underlying it, the nature of the dialogue to be undertaken with government and donor partners during design and implementation, management approaches (including those to support the Paris Principles of aid effectiveness), how conditionality will be

[54] Applicability to certain services problems for specialised services, which often require the full-time employment of staff and expensive equipment but have varying or low levels of demand. The top-up system included in the tariff for such treatments has not proved adequate. Mental health needs to be consensus on what the right approach is – significant variations in the way in which mental health services were provided; differing views about which treatments should be used for which conditions; and an absence of a strong causal relationship between interventions and outcomes. As a result the Department of Health has said that a national currency and tariff for mental health are 'still some way off'

established and how it will work in practice, how fiduciary risk will be identified, recorded and managed and how performance indicators will be established, monitored and evaluated. The design is oriented specifically towards the achievement of poverty reduction within an approach that explicitly aims to use and strengthen government systems to ensure sustainability and leverage (i.e. improving the effectiveness and pro-poor orientation of government spending). Evidence from budget support evaluations show that this investment in design and detail is reflected not only in terms of positive progress in the building blocks for poverty reduction (e.g. increased pro-poor spending, improved productivity of spending through strengthened PFM systems, etc.) but also in terms of minimising negative effects.

Whilst individual RBF and hybrid RBA/RBF schemes are not of the same scale as a typical budget support package (and, therefore, would not require the scale of design), budget support demonstrates the importance of good design and the amount of effort that it can take.

This underlines the importance of piloting, monitoring, evaluation, analysis and dissemination of results and of ensuring that capacity exists to carry these tasks out. This is the logic behind the establishment of GPOBA, which has been tasked with these responsibilities and, thereby, to establish the full potential of OBA schemes whilst ensuring that they will deliver results intended in an effective manner that minimises unintended effects. This will be particularly important in establishing a better understanding of the situations when OBA will work well, when it is not appropriate and the complementary interventions that will be required to get the best out of OBA schemes. It will also be important in establishing a better understanding of the extent to which aid delivery can (and should) be scaled up through OBA. It also underlines the importance of ensuring that the design of Cash on Delivery Aid is properly tested in practice before consideration is given to using this as a major vehicle to deliver aid.

Good design could help avert some unintended consequences. However, the complex nature of the interventions means that some unintended effects will be inevitable. This being the case the key will be to have monitoring systems in place which pick up these issues early and allow effective remedial actions to be put in place.

Box 6 presents a broad overview of findings on best practice from a recent World Bank review (though the finding would probably apply equally to RBA)

Box 6: Key lessons from RBF

- Political commitment and country ownership.
- Involvement of all relevant stakeholders in the design
- The need for a focused and gradual approach
- Adequate organizational structures and institutional capacity are key
- Need to focus on improving quality of services provided in addition to
- Increasing overall service provision and utilization.
- Selection of performance indicators is critical.
- Independent validation of achievement
- Adequate and appropriate monitoring and evaluation frameworks are critical.

Source: World Bank Taking Stock 2009

Some of the more specific key design issues are identified below – though the list is far from comprehensive:

- **Ensuring rewards are set high enough to elicit the necessary response**: In Uganda it was found that the small size of the bonus incentive on offer – between 5 to 7 percent of operating costs – was one of the main reasons why the RBF had no effect on the utilisation of services.[55]
- **The need to balance rewards with sanctions.** There is often an unwillingness to introduce and enforce sanctions yet many studies shown they can be effective in promoting better performance and improving value for money for funders or purchasers.[56] Eldridge and Palmer suggest that research should be carried out which applies the lessons from '1980s macro-level conditionality (particularly in terms of the lack of local ownership and failure to reduce funding when conditions not met)' to results based approaches to ascertain whether similar issues associated with lack of local control and lack of penalties applied could potentially become a feature of results based approaches and what responses may be appropriate.

- A broad message seems to be to **go for an approach which is relatively simple and is seen to be fair rather than try and identify the perfect approach**[57] …but, on the other hand, a simple approach may mean that easy to measure results are targeted rather than more relevant results and/or that there is a short-term focus at the extent of longer term sustainability.

- Ensure that a sound procurement process operates to the benefit of the scheme and to the cost and quality of output delivery. Evidence from review of OBA schemes, for example, demonstrates that **competition in procurement is important** and that it has resulted in some cases in lowering delivery cost below the level originally estimated in design.

- Demand risk may be substantial in some cases and may require **additional design effort** (e.g. use of NGOs or civil society to promote uptake of a service).

[55] The Vehicle Excise Duty concession available to Heavy Goods Vehicle (HGV) operators in possession of a Reduced Pollution Certificate is much smaller than the capital investment required to achieve the necessary emission standards. This has led to take-up of the scheme being low, with only around 15 per cent of HGVs in possession of a certificate. Prentice et al (2007),found awards of £50 to £150 paid in the Defence Aviation Repair Agency, and a bonus of £400 paid to all staff awarded an "exceeding" performance rating in the former Department for Constitutional Affairs were to small. Makinson (2000) recommended that performance-related pay bonuses should represent at least five per cent of base salary in order to be effective. (Rosenthal 2006) found that "maximum quality incentives average (d) 9% of plan payments (in a range of US P4P schemes), … but most physicians average less than a 5% bonus. He contrasted this to the UK QOF approach in which "physicians can more than double their income by achieving high scores on 149 quality indicators"

[56] Propper et al (2007) reviewed the hospital waiting time target scheme in England, which used the dismissal of key managers of hospitals as a sanction for poor performance. In comparison to Scotland – which had a similar focus on reducing hospital waiting times, but without the high-intensity sanction regime – there is evidence that the English scheme significantly reduced waiting times. Under P4P in the US some hospitals that score in the first and second deciles receive bonus payments from Medicare of two percent and one percent of Medicare payments for those services, respectively. After the first year, the demonstration used the bottom two deciles in each area of care to set baselines for poor performers. In the third and subsequent years, Medicare will reduce payments by up to two percentage points to hospitals that score below those baselines in the clinical areas involved. The Global Fund has also been able to release some $846m from low performing/non performing grants for reallocation to performing programmes. NGO have lost their service delivery contracts in Afghanistan.

[57] In the UK pragmatic approach was adopted with the changes phased in over time and adjustments made to create a "level playing field" Primary Care Trusts (PCTs were reimbursed for the difference between the reference costs of a local provider and the tariff price. This 'purchaser parity adjustment' payment was reduced year-on-year, from 50 per cent of the difference between reference costs and tariff in 2006/7 to 25 per cent in 2007/8, the final year in which it will be paid (Department of Health 2006f). A Market Forces Factor (MFF) has also been included to reflect the unavoidable additional costs associated with delivering services in certain parts of the country e.g. the higher unit costs of staff due to London.

- RBF and hybrid RBA/RBF design appears implicitly to assume that the requirement to demonstrate delivery of results will inevitably result in a strengthening of monitoring and reporting systems. Evidence suggests that this is not guaranteed, though, and further design effort may be required to address this issue.[58] Whilst schemes clearly do create a renewed opportunity to strengthen statistical and reporting systems, converting the opportunity into achievement that has benefits outside of the narrow focus of OBA reporting will require more in terms of dialogue and a considered approach than the simple existence of an OBA contract offers. The key will be to balance the independence of output verification with the broader sector monitoring needs, as well as to ensure ownership on the part of the relevant government agency. Similarly, it will be important to ensure that monitoring and review of schemes do deliver accountability and strengthen systems and do not just represent higher transaction costs.

- Design in some cases may need to be **more explicit with regard to aligning with government policies and strategies** for schemes to be sustainable and effective.

An overview of lessons learnt in relation to the individual schemes is shown in annex 11.

10. The case for conditionality needs to be continually revisited.

The effect of conditionality, as currently applied, on the achievement of results is far from clear. Where compliance (in terms of CCTs) was monitored in the Latin American schemes, it was found to be high (93-94%). However, positive impacts were also found in Ecuador on education enrolment despite no conditions being attached. In South Africa, a 'child support grant' had no conditions attached (i.e. an unconditional cash transfer) yet succeeded in improving child nutrition status if the family benefited when the child was below two years of age. This suggests that the full cost and infrastructure of a RBF arrangement may not be necessary in some cases to achieve compliance. Hence, there is a need to consider carefully whether it is worth the effort in making transfers conditional. The case for using the variable tranche of PRBS to encourage better performance is also open to question as referred to earlier

11. Fragile states pose challenges … but also offer significant opportunities

Although fragile states are not homogeneous they share a number of broad characteristics including:

- High levels of poverty: poverty is pervasive and at its most intractable in fragile states – and often poses a threat which goes beyond individual borders. Targeting is less of an issue when the majority of the population is poor.
- Lack of strong, legitimate Government leadership.
- Weak absorptive and implementation capacity (though there may well be significant non state capacity).
- Inefficient institutions with little legitimacy, undeveloped budgetary, accountability, monitoring and reporting systems.
- Lack of strategic frameworks such as PRSP or sector programmes.

[58] The November 2009 review of OBA Approaches notes that 'although OBA internalizes the monitoring of outputs, the monitoring framework established is rarely used for purposes other than payment of subsidies'. A case of 'what gets measured … gets done … and gets measured'.

- A large number of aid instruments and actors compared to other settings. Non aid instruments e.g. peace keeping play an important role. The key instruments used (e.g. project aid and humanitarian support) by their very nature make coordination much harder.
- Insecurity and political instability which are likely to disrupt aid flows and require higher levels of conditionality.
- High fiduciary risk.

In such circumstances it may well be difficult engage with Government let alone agree on what indicators should be used and what level of results are feasible. There is little capacity to track performance through administrative systems and broader surveys are carried out with less regularity than in more stable countries.

The extent to which results based schemes are feasible depends on donor attitudes to risk and perceptions of national commitment. Some risks are likely to be high – in particular, tracking progress and detecting misappropriation and gaming are likely to be much harder than in other countries. The trajectory of change should be as important as the existing situation. Taking risks is likely to be much more acceptable for donors in countries where capacity and commitment is growing even if this is from a very low base. Whilst achieving verifiable results may require greater up front investment in capacity building (in terms of both service delivery and statistical systems) the results themselves might actually cost less than in more developed countries. The GAVI ISS evaluation, for example, found the cost of immunising children to be much lower in less developed countries than in more developed ones which tend to have higher coverage levels. However, this is not necessarily the case across the board.

If donors wish to create a "level playing field" based on the principle that results based schemes should be "ambitious but challenging" for all countries this might imply that a common approach to defining results or setting targets in all countries might not be appropriate. Whilst fragile states might be expected to deliver comparable results to non-fragile states in the medium to long term, in the short term a mix of lower expectations about results and greater up front capacity building investment might be a more appropriate response.

Fragile states are likely to lack the leadership to proactively seek out RBA/RBF opportunities and may well be late adopters (as the GAVI ISS experience suggests) especially where application processes are complex. This being the case special measures – such as technical assistance or different (i.e. simpler) application processes for selected countries might make sense.

On a more positive note there is often significant non state capacity in fragile states and often political willingness to allow approaches such as contracting out which more stable countries might find politically sensitive. (In practice, with the collapse of public services there are fewer vested interests arguing against such service delivery models). Opportunities for innovation may therefore be higher in fragile states.

Key Messages

- RBA and RBF do deliver results but it is generally unclear whether this is due to the results focus or simply whether it is down to the extra money it brings. Well designed studies that could disentangle the two are rare. They are also difficult as most RBA/RBF schemes are implemented alongside a range of other reforms. As a result attribution is, and should be, extremely difficult. Whilst little is known about impact

even less is known about the cost effectiveness of RBA/RBF schemes (in relation to alternative approaches).

- Transaction costs are poorly defined and almost never measured. However, it seems clear that they are often extremely high (sometimes unnecessarily high) yet it is still unclear whether they achieve what they set out to deliver.
- Long term sustainability of the schemes is questionable. Agents may adapt, the method of implementation may weaken/fail to strengthen national systems and the schemes may reduce the need for necessary reforms.
- The equity picture is mixed and is driven as much by *who* implements as how the schemes are implemented. Targeting is possible but can be both complex and costly. Some design features may support more equitable outcomes.
- There are questions as to how applicable some RBF schemes are to low income countries.
- Politics matters. It is not only important that key policy makers support the schemes it also matters why they support them. Sound political analysis should be an integral part of any design phase.
- Good design is essential …. and may help to reduce unintended effects.
- The case for conditionality needs to be continually revisited. It is far from clear what it adds in terms of better performance in some settings.
- RBA/RBF schemes in fragile states involve significant risks but also offer major potential benefits. Judgments about the level of country commitment to delivering results (and taking the actions necessary to support this) will play a key role in assessing whether such risks are worth taking. Immediate results might not be forthcoming. Major up front investments may be needed to allow such states to respond to incentives introduced by the schemes. On a more positive note the subsequent cost of actually delivering results may be lower and opportunities for innovation might be greater than in stable settings as, in the case of the latter, political barriers may be lower. Special measures may be needed to allow fragile states access to RBA/RBF funding in the first place. Risks should be explicitly identified and managed.

7. Conclusions and lessons for DFID

The schemes are a means to an end – not ends in themselves. The ends need to be continually reviewed. DFID needs to continually take a step back and ask "are we buying the right results for the right people?" and "are these the results which will best accelerate progress towards the MDGs?" Only then should it consider which approaches it should adopt to achieve those results. If the answers to either of the above questions is "no", innovative RBA/F schemes may be, at best, an efficient way of heading towards the wrong outcome.

To adopt a supportive but cautious approach to innovative results based schemes. Whilst there is no shortage of discursive material associated with the implementation of RBA/RBF schemes the evidence base for its effectiveness generally remains weak not only in terms of what has been delivered by particular schemes, but on whether this delivery is economic, efficient and effective compared to alternative approaches. This review concludes that **whilst the evidence suggests that most RBA/F schemes do deliver results it is far from proven that these results are due to the results focus as opposed to the additional money**. Furthermore, it is far from clear where better results are achieved that the benefits associated with these outweigh the heavy, and often unnecessarily heavy, transactions costs associated with them (including the costs of targeting particular groups) or any negative impacts resulting from the neglect of other services not subject to financial incentives. It is therefore unclear, in the absence of further evaluation work, whether it would be appropriate to use any of the innovative RBA/F schemes as a means for substantially scaling up development assistance in line with international commitments.

The emphasis should be on **supporting innovation through well designed pilots with associated rigorous evaluation systems**. It strongly suggests, therefore, that any decision to use and/or scale up schemes may need to incorporate a research facility to demonstrate that the results obtained can at least to some degree be attributable to the aid resources delivered and that the scheme utilised is the most appropriate for the type of result to be achieved. (as indeed the HRITF has done). Whilst GPOBA in its capacity for collating and disseminating information about OBA, for example, provides an opportunity for this, evaluation material to date does not appear to have gone into this depth. (The Millennium Challenge Corporation, on the other hand, states that it does intend to carry out control evaluations, although it remains to be seen what the coverage and content of these will be.)

Budget support has a key role to play. Budget support stands out from the other schemes in that it can no longer be referred to as innovative.[59] It is tried and tested and is an integral component of the current aid architecture. The design and application of PRBS and EC MDG Contracts, for example, is based on decades of accumulated knowledge and experience (good and bad) of budget support delivered in a number of different forms, by a variety of institutions under some very different contractual and results based relationships – from the import support programmes of the 1980s operating under the now discredited top-down conditionality mechanisms of that time to the partnership approach of today. Much was learned during the 1980s and 1990s about what does not work in delivering aid through budget support. Much has been learned since about the current partnership approach (based on an evaluation of data and information from the late 1990s to the mid 2000s). Evaluation work supports the view that

[59] Although it should be noted that aspects of the EC MDG Contracts are innovative, particularly with regard to the length of the contract (6 years) and the performance nature of the variable tranche.

partnership budget support is facilitating the implementation of poverty reductions strategies and is facilitating the achievement of specific results (e.g. pro-poor expenditures, strengthened PFM systems) and also that it is doing this in a manner that is consistent with the Paris Principles for aid effectiveness. What the evidence is not yet able to show, though, is the impact of budget support on poverty per se – partly because of the complexities involved and partly because of the elapsed time that is required for this kind of impact to feed through properly into measurable data. This does not mean to say that budget support is without risk – fiduciary risk assessments demonstrate that risk is substantial in some cases. The design of the instrument, however, not only requires risk to be mapped and recorded, it requires risk management to become an integral component of budget support dialogue and management. Whilst budget support does not present opportunities for headline news in terms of the units of results bought the evidence shows that the performance framework it sits within is quietly delivering in most countries. The key question is how to make it work better in those countries where it is delivering and whether the package is working adequately in those countries where the results are more uncertain. In an environment where the UK is committed to expand its aid budget still further, it is also to be recognised, of course, that there will be a constraint in the extent to which budget support can be further scaled up,[60] and it will be important to identify and recognise the point at which this constraint is likely to bind.

There is certainly a place for the RBA/RBF schemes. There is undoubtedly a place for at least some RBA/RBF schemes in delivering aid and demonstrating results. There is also a clear demand for them – RBA and RBF have clear attractions, particularly in an environment when the public increasingly wants to know what their taxes actually buying. The question is "where?" Certain services lend themselves to such approaches but they need to be tailored to the country context. RBF and hybrid RBA/RBF schemes tend to be most appropriate where the required results are clear, where interventions are simple (and necessary incentives easily identified), where the provider has significant control and where capacity exists or latent capacity is available to respond to new incentives.

Complementary support will usually be required. In many cases capacity will need to be built up – only some of the schemes support this directly. In many (most??) countries the constraint might not be down to capacity – rather they may be due to inappropriate institutional arrangements. In such cases policy based support – perhaps backed up with some capacity building effort – might be the ideal approach.

Of the instruments reviewed here only budget support offers the potential to address many of these issues, with general budget support being more appropriate in some instances and sector budget support in others. General budget support tends to address important issues such as enhancing capacities for pro-poor policy formulation and budgeting, financial management, public sector reform, gender and equity issues at a macro level. Sector budget support, often underpinning a SWAp, is more appropriate for sector specific reforms and service delivery issues. Where a more holistic approach (e.g. through a budget support package) is required the concern is that results based funding or aid could be at best ineffective and at worst damaging.

Good design and preparation is essential. Clear understanding of the context and results chain. The context is extremely important. If diagnostic work suggests that there is latent capacity in place, that the results specified link appropriately to the impact desired and that financial incentives would work a pure results based approach might make sense. This is unlikely to be the case in many settings

[60] E.g. in terms of absorptive capacity among those countries where Partnership Principles are established and observed and possibly in terms of managing fiduciary risk in some cases.

If DFID sees RBA/RBF as a simple way of attributing development results to its assistance it will be disappointed. Attribution is, and *should* be, next to impossible for a variety of reasons. Firstly, it is currently extremely difficult to know which results are due to additional funding and which are due to the range of schemes being employed to improve performance. Secondly, because RBA/RBF is typically implemented as part of an overall package we don't know which mechanisms should take credit for these results. A key challenge will be ensuring the donors can coordinate their activities to ensure that a balanced package of support is provided. In short we don't know which results are additional and of those that are (if any) we don't know who to attribute them to. Better evaluation might help with the first – the secondly is simply a feature of the complexity of system strengthening. .

DFID should try and seek synergies between RBA/RBF schemes and other financing initiatives… and avoid competition and fragmentation. Although innovative RBF and hybrid RBA/RBF approaches may result in short term gains there is a significant risk that in bypassing Government systems some of them will serve to undermine the development of national systems. In doing so they may reduce the value, possibly substantially, of the results obtained. Whilst OBA schemes, for example, require government approval, there is nothing inherent in their design that makes them more likely to operate any differently to traditional projects in terms of aligning with government policies and strategies. Whilst they remain a relatively minor component of service delivery, this is unlikely to be a major problem. This may not remain the case if aid is scaled up through OBA approaches.

In the health sector, there is a further risk that support – especially that provided through global initiatives – will serve to fragment approaches at the country level. The worst case scenario is one in which a second wave of global health partnerships focusing not on specific diseases – but various components within the existing health system strengthening building blocks – further fragments support. Within the health financing building block RBF is competing for space with Providing for Health which focuses heavily of social protection and social insurance as well as the soon to be established Centre for Progressive Health Financing (CPHF) which has a broad mandate to support pro poor financing. One senior policy maker closely involved in this debate suggested that an important role of the CPHF might be to help build the evidence base for results based approaches The Joint Funding Platform which is attempting to develop country level mechanisms to channel resources from a range of global initiatives also clearly has an important role to play in the health sector.

DFID needs consider the extent to which it is willing to trade off adherence to the Paris Principles with the need for short term results. A potentially very important issue that the literature does not seem to address is the impact of innovative RBA and RBF schemes on the Paris Principles and Accra agenda. Whilst budget support is explicitly designed to be consistent with Paris and Accra, the same cannot necessarily be said of RBF and hybrid RBA/RBF schemes which typically by-pass government systems completely. To the extent that by-passing government systems represents an effective contracting out of service delivery by the government, with participation and ownership of the process by the government this matters little. However, it is not self evident that this is the case. There is a risk that the service delivery in question will revert to the public sector, with an eroded capacity to manage and implement and/or that the public sector is left with a model of contracted-out delivery and has not developed a capacity to procure, manage, monitor and verify. Notwithstanding formal requirements to obtain government sanction to proceed with such schemes (on a no objection basis, for example, in the case of World Bank administered OBA schemes), there is a real risk that they will effectively be conceived and implemented externally and completely by-pass government policy, planning and PFM systems. Fragmenting policy decisions and policy implementation is a

potentially high price to pay to demonstrate the achievement of (a sub-set of) sector results in this way. This suggests that a decision to use and or scale up the use of RBF and hybrid RBA/RBF schemes should explicitly incorporate procedures to ensure dialogue with and ownership by government.

Developing administrative government systems and auditing them is preferable, in principle, but there are real practical obstacles to overcome. Data in developing countries is often collected through parallel systems. It is generally possible to control quality – though this also needs work. However, parallel systems may be expensive and setting them up can represent a missed opportunity to develop government systems and may actually undermine existing systems This raises the question as to whether enhanced, audited government systems can be used to meet the need for data under RBA/RBF schemes. There are currently few examples of auditing in government education systems. There would be a greater need to audit government systems and a need to do so more carefully. Data triangulation with other sources – usually household surveys – is often used to provide a broad reality check, but this would be too weak for RBA/RBF schemes. This is therefore an area that would require considerable development and where there is little experience to draw on. There is clearly the potential for conflict where audits uncover problems.

Imprecision and inaccuracies in survey data may cause concern. There are always sampling errors, which are quantified on the DHS and some others. These are not usually great at national level for the indicators used, though this may not hold for some (e.g. maternal mortality). Second, estimates may be biased. For example, Carr-Hill estimates the effects of omitting the non-household population on the out-of-school population to be quite high. The focus on change, rather than level, and the general consistency of the survey basis reduces the effect of this.

DFID needs to ensure that equity remains a key objective of the schemes. There is a risk that RBA and RBF schemes might deliver demonstrable results, but at the cost of disadvantaging some low income countries. The GAVI ISS evaluation finding that LICUS countries “tended to apply for GAVI ISS funding much later than other countries, (…. and) were also less likely to receive reward funding” demonstrates this risk. Similarly, one of the concerns over the concept of CODA aid is that fragile countries and countries with lower capacities generally (usually the most needy) would be less likely to respond well to the incentive payments in getting children in primary school and retaining them there. This suggests that if CODA is to be used to scale up aid, it may need to be complemented with more direct interventions is weaker capacity systems.

For DFID embracing RBF and hybrid RBA/RBF approaches will entail a shift of risks way from fiduciary risks to issues related to data quality and effectiveness of monitoring systems.

The need for dedicated support to build statistical capacity will remain – RBA/F will not solve this. Whilst an oft claimed advantage of RBA and RBF schemes is that it provides incentives for government to strengthen statistical and reporting systems, there is some evidence that suggests the focus remains on the performance measures concerned. This suggests that an automatic strengthening of statistical and reporting systems over time should not necessarily be assumed. DFID and other partners need to renew their efforts to encourage and support statistical capacity building. The new Statistics for Results Facility may play an important role here.

Key messages

- Need to focus on the bigger picture. Important to ask “have the right results have been chosen?” and “for the right people?” before asking “does the scheme actually delivers results?”
- A good diagnosis of the underlying problems and main constraint to scaling up is essential.
- Identifying the right results and understanding the complexities of the results chain and the location of these results within it is essential.
- DFID should not only strongly encourage piloting and testing of approaches but should insist that innovative schemes have appropriate evaluation arrangements – to date most have not incorporated a control element.
- The costs of moving away from Paris Principles in return for making results appear more transparent should be documented clearly and acknowledged.
- Pure RBF/A is only likely to be appropriate in some settings. Many of the schemes build on a range of approaches and do not rely solely on results based financing/aid are likely to be applicable in a wider range of circumstances. Other schemes which rely heavily or exclusively on funding performance may need to be complemented by other mechanisms in some settings. Attribution may be easier to detect for stand alone RBA/F approaches; however, such approaches are probably less likely to be successful.
- PRBS has the strongest and, generally positive, evidence base and supports the Paris Principles. It should be the instrument of choice and used unless there are compelling reasons not to. The challenge is to make it work better. For RBF, by contrast, the challenge it is to demonstrate the circumstances in which it is worthwhile.
- Attribution is possible in a very narrow sense only and may not relate well to the desired developmental impact. If RBA/RBF is appropriately implemented alongside a well thought through, balanced package of services, attribution will probably never be possible.
- DFID should look for synergies between global financing initiatives RBA and RBF. The CPHF offers some good opportunities.

Key references

Abt Associates (2007) Evaluation of the first five years of GAVI ISS.

American Medical Association (2005) Principles for Pay-For-Performance Programs. June 21.

Audit Commission (2005) Early Lessons from Payment by Results.

Audit Commission (2008) The right result? Payment by Results 2003-07.

Bhatia MR and Gorter AC (2007). Improving access to reproductive and child health services in developing countries: are competitive voucher schemes an option? Journal of International Development Vol 19, Issue: 7975-981.

Booth & Lucas (2001) Desk Study of Good Practice in the Development of PRSP Indicators and Monitoring Systems, Final Report, DFID, November 3, 2001

Brenzel, L (2009) Taking Stock: World Bank Experience with Results-Based Financing (RBF) for Health. World Bank.

Carr-Hill, Hopkins, Riddell and Lintott (1999) Monitoring the performance of education programmes in developing countries, DFID Education Research, Serial number 37.

Carr-Hill, R (2009) Issues in Data and Monitoring and Evaluation, Mid-Term Evaluation of the EFA Fast Track Initiative Working Paper 3[a].

Chapman, K. (2006) Using Social Transfers to scale up Equitable Access to Education and Health Services. Background Paper for DFID.

Chee, G et al (2007) Evaluation of the First Five Years' of GAVI Immunization Services Support Funding. Abt Associates Inc.

Cheung, S (1992) On the New Institutional Economics. In: Werin and H. Wijkander (eds.) Contract Economics. Blackwell Publishers, pp. 48-65

Crouch and Mitchell (2008) Audit Options to Certify Results for a Cash on Delivery Contract in the Education Sector, by Luis Crouch and Jonathan Mitchell (RTI International), Version 2, June 30.

Demsetz, H (2003) Ownership and the Externality Problem. In Anderson and McChesney (eds.) Property Rights: Cooperation, Conflict, and Law. Princeton University Press.

Department of Health (2002) Reforming NHS financial flows: introducing payment by results.

DFID (2006) Guidance on Aid Instruments: A DFID Practice Paper.

DFID (2006) Using Social Transfers to improve human development. Social Protection Briefing Note 3.

DFID (undated) Experience of voucher schemes in health.

Eldridge, C and Palmer N (2009) Performance-based payment: some reflections on the discourse, evidence and unanswered questions. Health Policy and Planning 24 (3):160–166.

Ellison, R (2009) The statistical monitoring of Sarva Shiksha Abhiyan (SSA). Technical Paper, 1st draft, October 2009]

Ensor, T (2003) Consumer-led demand side financing for health and education: an international review.

Epstein, AM (2007). Pay for Performance at the Tipping Point. The New England Journal of Medicine 356 (3): 515–517.

ESSPIN (2009) Education Sector Support Programme in Nigeria (ESSPIN) Approach to supporting EMIS.

Farrar et al (2009) Has payment by results affected the way that English hospitals provide care? Difference-in-differences. BMJ 339 doi: 10.1136/bmj.b3047

Gannicott, K (2008) Kwara State ESA [Education Sector Analysis], Second draft, March 2008.

Ghana Statistical Service, Ghana Health Service and ICF Macro (2009) Ghana DHS Report, 2005 Demographic and Health Survey 2008, by Ghana Accra, Ghana: GSS, GHS, and ICF Macro. Available from DHS Macro International Website

Ghana Statistical Service, Ghana Health Service and ICF Macro (2008) Demographic and Health Survey 2008.

Glassman, Gaarder and Todd (2006) Demand Side Incentives for Better Health for the Poor: CCT programs in Latin America and the Caribbean. IADB.

Glickman SW et al (2007) Pay for Performance, Quality of Care, and Outcomes in Acute Myocardial Infarction JAMA. Jun 6; 297(21):2373-80.

Global Fund to Fight AIDS, TB and Malaria (2009) Performance-based Funding at the Global Fund. Power point presentation.

Global Fund to Fight AIDS, TB and Malaria (2008) Guidelines for Proposals – Round 9 (Single Country Applicants).

Gorter A. et al (2003) Competitive Voucher Schemes for Health: Background Paper. ICAS.

Governance and Social Development Resource Centre (2006) Desk based review on experiences in demand side financing for maternal health.

Hanson, KN et al (2003) Cost-effectiveness of social marketing of insecticide-treated nets for malaria control in the United Republic of Tanzania. Bulletin of the World Health Organisation, 81: 269-276.

Hatt, L et al (2010). Economic Evaluation of Demand-Side Financing (DSF) for Maternal Health in Bangladesh [Draft]. Review, Analysis and Assessment of Issues Related to Health Care Financing and Health Economics in Bangladesh, Abt Associates Inc.

HLSP (2008) Health sector monitoring: Approaches, issues and lessons from a review of eleven countries.

HLSP (2009) GAVI Health Systems Strengthening Support Evaluation.

Hyun H. Son (2008) Conditional Cash Transfer Programs: An Effective Tool for Poverty Alleviation?

IDD and Associates (2006) Evaluation of General Budget Support: Synthesis Report.

Ir, P, Horeman D., Narin S., and Van Damme, W (2008). Improving access to safe delivery for poor pregnant women: a case study of vouchers plus health equity funds in three health districts in Cambodia. Studies in HSO & P. 24:225-256.

Kaushal, S (2009) Post Enumeration Survey of DISE data: major findings 2007-08 – a compilation. NUEPA.

Kenya Voucher scheme website. http://www.output-based-aid.net/e102/

Koehlmoos TLP et al. (2008). Rapid assessment of demand-side financing experiences in Bangladesh. Working Paper 170. International Centre for Diarrhoeal Disease Research, Bangladesh.

Kundu, F (2007) A new initiative towards universal access to reproductive health in Kenya. National Coordinating Agency for Population & Development.

Lagarde, Haines and Palmer (2007). Conditional Cash Tranfers for improving uptake of health interventions in Low and Middle Income Countries: A systematic review. JAMA, 298, 16.

Lenel, A, Griffith, D, Anschutz K (2009) Voucher schemes in the health sector: The experience of German Financial Cooperation. KfW Entwicklungsbank.

Levitt and Dubner, 2005, Freakonomics: A Rogue Economist Explores the Hidden Side of Everything.

Loevinsohn, B (2008) Performance-Based Contracting for Health Services in Developing Countries: A Toolkit. World Bank 2008

Lu C, Michaud CM, Gakidou E, Khan K, Murray CJ. (2006) Effect of the Global Alliance for Vaccines and Immunisation on diphtheria, tetanus, and pertussis vaccine coverage: an independent assessment. Lancet. 368(9541):1088-95.

Luci, B. Cash incentives for the poor (2007). Gotham Gazette.

Macro International (2009) The Five-Year Evaluation of the Global Fund to Fight AIDS, Tuberculosis, and Malaria: Synthesis of Study Areas 1, 2 and 3.

Marchant, T et al (2002) Socially marketed insecticide-treated nets improve malaria and anaemia in pregnancy in southern Tanzania. Tropical Medicine & International Health, 7(2): 149-158.

Maybin, J (2007) Payment by Results. King's Fund Briefing.

Mehta, A. (2008) Analytical Report 2006-07: Elementary Education – progress towards UEE. UEPA. http://www.dise.in//ar2005.html (see Analytical Report 2006-07; section on 'School and facility-related indicators')

Mellstrom C, Johannson M (2008) Crowding Out in Blood Donation: Was Titmuss Right? Journal of the European Economic Association (RSS) doi: 10.1162/JEEA.2008.6.4.845

Meredith B et al (2005) Early Experience with Pay-for-Performance: From Concept to Practice (abstract). JAMA 294 (14): 1788–1793.

Meuwissen LE et al (2006) Does a competitive voucher program for adolescents improve the quality of reproductive health care? A simulated patient study in Nicaragua. BMC Public Health, 6: 204.

Milazzo, A (2009) Conditional Cash Transfers: An annotated bibliography.

National Audit Office (2008) Department for International Development: providing budget support to developing countries, 8 February 2008.

National Audit Office (2008) NHS Pay Modernisation: New Contracts for General Practice Services in England.

National Audit Office (2008) The use of sanctions and rewards in the public sector.

Olsen, IT (2009) Result-based Financing in the Health Sector: Experiences from Norway and from low-income countries.

Pearson, S et al (20080 The Impact of Pay-for-Performance on Health Care Quality in Massachusetts, 2001-2003. Health Affairs, 27, no. 4 (2008): 1167-1176.

Petersen et al (2006) Does Pay-for-Performance Improve the Quality of Health Care? Petersen, Woodard, Urech, Daw and Sookanan, Annals of Internal Medicine 2006 (abstract)

Regalía F, Castro, L (2007) Performance Based Incentives for Health: Demand and Supply Incentives in the Nicaraguan Red de Protección Social. CGD Working Paper 119

Rosenthal MB and Frank RG (2006) What Is the Empirical Basis for Paying for Quality in Health Care? Medical Care Research and Review 63 (2): 135–57.

Rusa, L. et al (2009) Rwanda: Performance-Based Financing in the Public Sector. Center for Global Development.

Smith, P (1993) Outcome-related performance indicators and organisational control in the Public Sector, British Journal of Management, Vol 4, pp135-51

Standing, H (2004) Understanding the demand side in service delivery: definitions, frameworks and tools from the health sector. DFID Health Resource Centre.

Steele N, Maisey S, Clark A, et al. (2007) Quality of clinical primary care and targeted incentive payments: an observational study. Br J Gen Pract. 2007; 57:449–454.

Strategic Partnership with Africa (2005) Sector Budget Support: A Note from the Dublin Workshop of SPA Working Groups, 5-6 October 2005.

Sussex J, Street A. (2005) Activity-based financing for hospitals: English policy and international experience. Office of Health Economics.

Toonen et al (2009) Learning Lessons on Implementing Performance Based Financing from a Multi Country Evaluation. KIT (Royal Tropical Institute).

UNSD (2003) Indicators for Monitoring the Millennium Development Goals: Definitions Rationale Concepts and Sources, ST/ESA/STAT/SER.F/95, UNDG.

Vaillancourt (2009) Do Health sector Wide Approaches Achieve Results? Emerging Evidence and Lessons from Six Countries. World Bank.

Walford, V (2008) Demand side incentives in health – learning from the Asia experience. Report for DFID.

Annex 1: Template for analysis of approaches

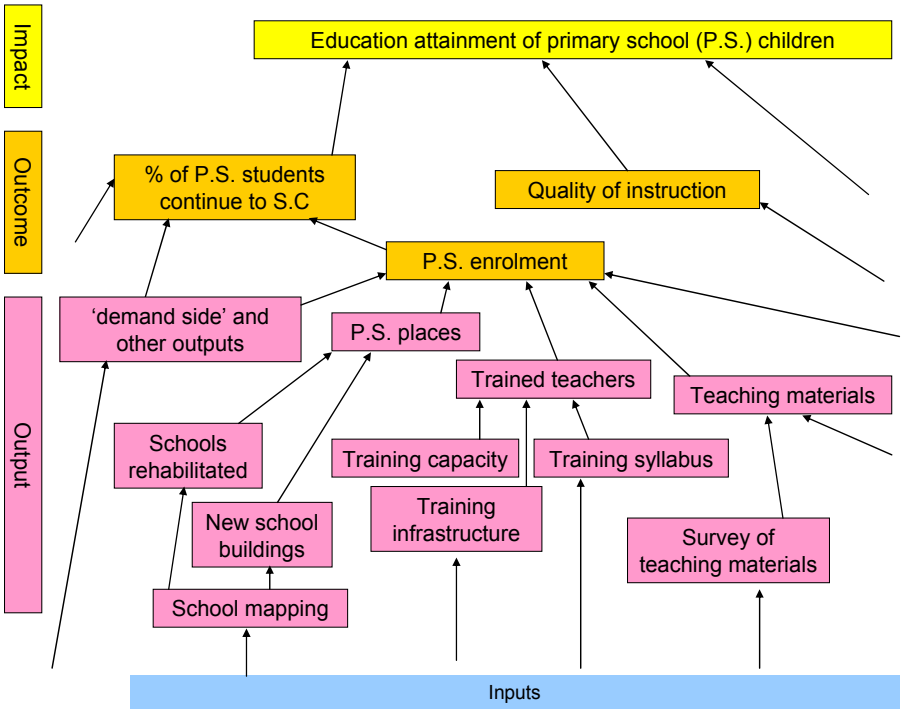
1. Approach /Scheme	**Name of scheme plus short description.**
2. Type of scheme	– Project team view of what the scheme should, in fact, be categorised as (in terms of a standardised list of scheme types that we will propose as part of the overview report)
3. Design of scheme, including performance and link to reward mechanism	– Nature of dialogue with beneficiary with regard to: • design of scheme and associated payment / reward mechanism • implementation of scheme and associated payment / reward mechanism – Main factors taken into account in establishing • type and/or level of performance • level of reward / payment. – Description of how the scheme defines performance – Relationship between performance achieved and disbursal of funds – Component of payment that is 'core' and component that is paid against performance – Also to include analysis and comment on: • factors determining allocation between beneficiaries when the initial recipient is not the final beneficiary • extent of earmarking of payment / reward
4. Conditionality	– Nature of conditionality • Ex post or ex ante conditions? • When are payments made – up front? in tranches? at end of the contract or review period? • Is over performance rewarded?
5. How the "performance contract" is enforced?	– Legal contract? – Dialogue and MoU?
6. Evidence of impact/analysis of schemes	– Achievement of performance targets? – Evidence of impact beyond immediate performance targets (if any and if requisite data available)? – Any unintended impact – e.g. through perverse incentives? – Consistency with Paris Principles (e.g. Ownership, alignment, mutual accountability, predictability, support for country systems)? – likely equity impacts: selection/targeting of beneficiaries; – efficiency considerations: associated costs (including transaction costs
7. Indicators	– Which indicators trigger payment? – How many indicators? – What are they? – What level (output / outcome etc.)? – Are they good proxy measures? – What evidence or scope is there for gaming and perverse incentives? – Are poverty and equity explicitly considered? – Are there clear payment rules on achievement of some but not necessarily all indicators?
8. Data	– Which data sources are used? – Frequency? – Time lag? – Quality (coverage, accuracy, non response & response bias)?

	– Who collects? – Independent collection or verification?
9. Key measurement issues	– Main problem(s) or gap(s) with indicators and data
10. Possible measurement solutions	– Ways to address main problems and gaps for consideration, preferably drawing on existing sources and without undermining national statistics systems.

Annex 2: Complexities of the results chain/tree

The results chain is actually more complex than its normal linear presentation and can be thought of usefully as a results tree. Presenting the 'result chain' in the form of a tree demonstrates more clearly that it is possible to choose outputs in a results framework that are relatively close to the desired outcome or relatively further away. It also demonstrates the potential complexity of ensuring that results are specified, planned and budgeted across a range of outputs to ensure the delivery of a particular level and quality for an outcome (and thereby emphasises the importance of alignment with regard to use of RBF schemes in particular).

In the following diagram for the education sector, for example, which shows just some of the outputs and outcomes associated with improving educational attainment, focussing on rehabilitating primary schools and/or establishing new primary schools in a results framework may have a very different effect on enrolment if this is accompanied by measures to enhance demand (e.g. by removing user fees, providing school transport, etc.) than if demand issues are left unaddressed. Similarly, if the availability of trained teachers, instruction materials and other outputs is neglected, this will make higher enrolment more difficult to achieve. A focus on enrolment, therefore, effectively assumes that capacity exists to deliver a range of outputs to make enhanced enrolment possible and / or that the use of any RBA/RBF schemes aimed at the delivery of one or more outputs in this tree are sufficiently aligned with sector policy, planning and budgeting to ensure that higher enrolment is achieved as planned. In terms of its impact on educational attainment, however, focussing on the single outcome of 'enrolment' may be insufficient if other outcomes, including those related to quality of instruction, are not also addressed.



Similarly, if rehabilitation of school buildings is chosen as the output to focus on (e.g. because this easier to control and deliver in a RBF framework), whilst it may be possible to commission delivery of this output in a performance framework with a high likelihood of achieving the target specified, this would have a much weaker link to the intended impact than would a focus on enrolment.

Annex 3: DAC definitions

DAC Definitions

Impacts
Positive and negative, **primary and secondary long-term effects** produced by a development intervention, directly or indirectly, intended or unintended

Outcome
The likely or achieved **short-term and medium-term effects** of an intervention's outputs. Related terms: result, outputs, impacts, effect.

Outputs
The **products, capital goods and services** which result from a development intervention; may also include changes resulting from the intervention which are relevant to the achievement of outcomes

Results
The **output, outcome or impact** (intended or unintended, positive and/or negative) of a development intervention. Related terms: outcome, effect, impacts

Results Chain
The causal sequence for a development intervention that stipulates the **necessary sequence** to achieve desired objectives beginning with **inputs, moving through activities and outputs, and culminating in outcomes, impacts**, and feedback. In some agencies, reach is part of the results chain. Related terms: assumptions, results framework.

Results framework
The **program logic** that explains how the development objective is to be achieved, including **causal relationships and underlying assumptions**. Related terms: results chain, logical framework.

Performance
The **degree to which** a development intervention or a development partner operates according to specific criteria/standards/ guidelines or **achieves results in accordance with stated goals or plans**.

Performance indicator
A variable that **allows the verification of changes in the development intervention or shows results** relative to what was planned. Related terms: performance monitoring, performance measurement

Performance measurement
A system for assessing performance of development interventions **against stated goals**. Related terms: performance monitoring, Indicator

Goal.
The **higher-order objective** to which a development intervention is intended to contribute. Related term: development objective

Annex 4: Which performance levers do the schemes use?

Scheme and Lavers	Funding	Regulation and Standards	Targets	Taxation	Legislation	Partnerships and MoU	Contracts	Performance and Staff Management	Competition and Choice	Education	Publishing Performance Data
GFATM											
GAVI ISS											
UK QOF											
UK PbR											
US P4P											
Vouchers											
CCTs											
PRBS											
EC MDGc											
CODA											
Millennium Challenge Account											

Annex 5: Key design features

	GFATM	GAVI ISS	Vouchers	CCTs	UK QOF	UK PbR	US P4P
Degree of earmarking	To the three diseases	In principle – unearmarked	To targeted individuals (in kind)	To targeted individuals	To GPs for the delivery of certain outputs	To providers for the delivery of certain outputs	
Use of national systems	Rarely. Occasional participation in SWAps/pooled funding. Sometimes on plan. Parallel accountability arrangements. No engagement in sector reviews/MTEF processes	Usually uses Government systems and on plan No engagement in sector reviews/MTEF processes	Rarely. Maternal voucher scheme supported through parallel funds in the SWAp	Yes. Many – in middle income countries are domestically funded	Yes	Yes	Yes
Level of performance	Broad	Health services delivered (focus on quantity)	Health services (and associated services e.g. transport) focus on quantity)	Health service outputs (focus on quantity	Health services outputs. Focus on quality issues Process indicators related to management – patients views of quality of care	Health service outputs (focus on quantity)	Health services outputs. Focus on quality issues
Definition of Performance	Set out in proposal (could be volume of services or performance against targets) Limited number of outputs and outcome indicators. Set at the country level – ideally taken	Volume: Number of children immunised against DPT 3	Volume: Delivery of specified service (often by accredited provider	Volume: Utilisation of specified service by intended beneficiary	Targets: Range of ~ 135 indicators	Volume: Delivery of a services according to an HRG	???

	from existing country frameworks						
Who sets performance goals?	Agreement between country and GFATM Set out in proposal	No goals as such – basically a fee for service approach	Funding agency	Funding agency	Dept of Health	No specific goals – basically fee for service approach	Purchaser (Medicare)
Type of Reward	Financial – but also preferential access to additional resources (Rolling Continuation Channel). Sanctions: funds not disbursed – no access to RCC	Financial – for MoH. Sanctions: no performance – no reward	In kind for beneficiary. Financial for service provider	Financial for beneficiary	Financial for GPs Sanctions: no performance – no reward. Indicators have been modified over time (made more challenging)	Financial for provider Sanctions: no performance – no reward	Financial for provider. Providers compete with each other to get rewards. Sanctions in apply for poor performers (reduced FFS payments
Limit to rewards	According to proposal budget	Limited by number of children – only 100% can be immunised	Distribution is typically targeted to specific target groups e.g. the poor, women and children or geographically	Covered typically directed to specific target groups e.g. the poor, women and children or geographically	Maximum of 1000 points available	In theory few limits for an individual provider. In practice limited by purchasers budget/ preference to purchase services locally	
Validation of performance	Local Funding Agent	Data Quality Audit by external audit company			Trust basis. Limited audit by PCTs		
Beneficiary of reward	Principal Recipient – MoH or NGO body	MoH	Provider receives reward – voucher recipient receives service	Individual	General Practitioner	Provider (Hospital)	Provider
Level of Reward	Agreed in Proposal	$20 per child immunised against DPT 3 above self reported baseline	Voucher covers cost of delivering a specified service. May also cover indirect costs e.g. travel costs		£125 per point		

Type of reward	Depends on Board recommendation	Directly related to level of performance	All or nothing	All or nothing	Directly related to level of performance between lower and upper thresholds	Directly related to level of performance	
Basis for level of reward	Budget as agreed in proposal	Estimated cost of immunising a child	Negotiated with eligible providers	Set by funding agency according to the level they think necessary to encourage people to utilise services	Set by DoH based on assumptions about likely performance and available funds. No specific rationale for level	National average cost of providing service	
Award of Reward	Administrative decision made on recommendation of LFA. Based on a comprehensive overview of performance not just against indicators Also includes financial performance and quality of grant management	Automatic based on verified data	In advance to target groups (it is up to the recipient to redeem)	Automatic after service has been used/action taken	Automatic after performance estimated and verified	Automatic after acceptable service delivered	
Flexibility on what rewards can be spent on	None – spent according to proposal (unless changes agreed)	Full flexibility	Voucher must be used for specified service. Provider can do what they like with the revenue	Full flexibility. Recipient can do what they like with the money (though they are encouraged often with health education etc to use it for particular purposes	Reinvested in facility services	Reinvested in facility services	
Share of recipients	First two years of proposal represents	Initial ISS support represents core			Modest – 1/3	High	

income	core grant – last three years are performance based. Share between them is variable Proposal is often frontloaded	funding – reward payments kick in in later years. Relative balance between them depends on performance					
Conditions		Minimum 80% performance in DQA					

	CODA	**GPOBA**	**PRBS**	**MDG Contracts**	**MCA**
Degree of earmarking	Unearmarked	In principle – unearmarked but estimated cost of subsidy required for output	Unearmarked (although SBS earmarked to sector)	Unearmarked	Earmarked to MCA schemes
Use of national systems	Yes	No	Yes	Yes	No
Level of performance	Narrow (primary education completion)	Narrow (output delivery)	Broad (poverty reduction; improving PFM, promoting good governance and transparency; fighting corruption	Broad (national policy and strategy; macroeconomic stability; PFM improvement)	Broad (6 'ruling justly' indicators; 6 'encouraging economic freedom' indicators; 5 'investing in people' indicators)
Definition of Performance	Primary education completion above a given baseline	Outputs delivered as contractually agreed	'Satisfactory progress' against Partnership Criteria (fixed tranche); achievement of specific targets from PAF (variable tranche)	'Positive trajectory' against Eligibility Criteria (fixed tranche); achievement of specific targets from PAF (variable tranches)	performance compared to peers with regard to 17 indicators
Who sets performance goals?	Established in open contract by donor.	Established in open contract by donor.	Donor and government in partnership through dialogue	Donor and government in partnership through dialogue	Donor chooses indicators and decision methodology
Type of Reward	Financial	Financial	Financial – but package often includes provision for TA	Financial – but package often includes provision for TA	Financial
Limit to rewards	Limited by numbers completing primary education	Limited by number of outputs delivered	Limited by agreed size of budget support	Limited by agreed size of budget support	Limited by cost of projects proposed and agreed to be funded
Validation of	To be determined	Independent verification	Joint donor / government	Joint donor / government	Independent verification

performance		agent	assessment of performance	assessment of performance	
Beneficiary of reward	To be determined, but probably through budget process	Implementing agent	Various government agencies through budget process	Various government agencies through budget process	Project implementers (through government's MCA accountable entity)
Level of Reward	To be agreed in contract	Subsidy for delivery of output (agreed in contract)	determined by fiscal framework parameters, including fiscal deficit policy and other donor grant receipts	determined by fiscal framework parameters, including fiscal deficit policy and other donor grant receipts	Defined by country proposal and MCC response to this
Type of reward	Directly proportionate to primary school completers above baseline	Directly related to level of performance	All or nothing for fixed tranche (although fixed tranche can also be reduced if necessary). In proportion to performance for variable tranche	All or nothing for fixed tranche. In proportion to performance for variable tranches	All or nothing
Basis for level of reward	Amount per child as agreed in contract	Estimated cost of required subsidy	Agreed in dialogue between DFID and government	Agreed in dialogue between DFID and government	Cost of projects proposed by government
Award of Reward	Automatic following independent verification of delivery	Automatic following independent verification of delivery	In advance and according to agreed timetable (fixed component). Following verification of results (variable tranches)	In advance and according to agreed timetable (fixed component). Following verification of results (variable tranches)	In advance and according to agreed timetable
Flexibility on what rewards can be spent on	Full flexibility	Full flexibility	Full flexibility (although with some sector earmarking for SBS)	Full flexibility	To be spent on MCA projects as proposed by government
Share of recipients income					
Conditions	Delivery of improved primary school completion	Delivery of outputs	Satisfactory progress with regard to Partnership Principles (fixed and variable tranches). Delivery of agreed results (variable tranche)	Positive trajectory with regard to Eligibility Criteria (fixed and variable tranches). Delivery of agreed results (variable	Eligibility may be suspended or terminated for: significant policy reversal; or pattern of actions inconsistent with the

				tranches)	eligibility criteria

Annex 6: Overview of schemes and their implications

Design Feature	Examples	Implications (for incentives)
Cash or in kind payment or other benefits and degree of flexibility with which funds can be used	Vouchers provide in kind support/GAVI ISS provides $20 per additional child immunised and there is flexibility on how rewards can be used. For the GFATM good performance provides access to Rolling Continuation channel. Reward funds must be used for agreed purposes	Flexibility in use of funds is more consistent with country ownership. Potential for funds to be used more efficiently (in line with local needs) but also potential for misuse (could argue countries have earned the right to use the money however they want and that this is irrelevant)
Targets can be set at different levels – e.g. outcome, output, input level …. Or at global level, national level of local level	GFATM uses a range of country level indicators at all level – input, outputs, and outcomes. PbR/GAVI ISS focus on service outputs.	The higher the level the more difficult attribution becomes The incentive for the agent is to negotiate as low a target as possible rather than one which is challenging but achievable
Focus on quantity v focus on quality	GAVI ISS focuses on number of children immunised. QOF/US P4P focuses on quality	Focus on quantity raises risk than quality will be compromised. Measuring quality is extremely challenging – heavy reliance on subjective judgement
Ex post or ex ante conditions	GFATM – demonstrated good performance opens access to Rolling Continuation Channel	
Administrative or automatic decisions over release of funds	GFATM performance grants (years 3-5) subjective to Board approval. PbR payments automatic	Transactions costs involved in administrative process/uncertainty over results
All or nothing v graduated payments	QOF offers graduated payments above a minimum threshold	If providers think they are unlikely to meet targets they may not try at all
Transitional measures		
Target or just services provided. If targets they can be constant or standards increasing to help drive performance	QOF sets targets. GAVI ISS focuses on number of children immunised	Potential distortions e.g. for GAVI ISS countries with
Who sets targets	GAVI ISS: countries set targets. QOF targets set by Department of Health	Some question under QOF that national targets might be unfair and that locally set targets might be more appropriate
Degree of reliance on independent validation	GAVI ISS and GFATM rely on external Data Quality Audit and Local Funding Agent report. QOF: high trust model relies on self reporting/limited audit by Primary Care Trust	High trust model implies lower transactions cost but more potential for gaming/failure to uncover and deal with gaming
Overall allocation – formula based e.g. GAVI – proposal led GFATM – implications for equity	PbR and QOF – overall resource envelope is based on a needs based resource allocation formula/ GFATM is based on proposal	Needs based formula can ensure equity in geographical inputs (if not outputs and outcomes). Proposal based approaches may compromise equity if based on country assessed financing gap

		(GFATM) or reliant on country capacity to apply (GAVI ISS)
Scope of services:	Rwanda broad range of services are covered In Tanzania it is narrow	Where range is narrow there is a risk that services which do not receive particular incentives are ignored
Different limits on rewards	For vouchers/CCTs each beneficiary is entitled to one reward. Under PbR providers can increase the number of rewards they have access to	Potential to earn additional income provides strong incentive but can also create uncertainty in terms of long term planning. Under PbR the impact of financial volatility associated with the payment scheme was a problem
Degree of competition for reward	QoF – no competition – if the standard is reached the reward will be given. In US P4P/PbR there is competition between providers – there is no standard but the lowest performers do not get reward or get a lower reimbursement. In the case of MCA countries compete to have access to potential funds on the basis of past performance	Under QOF the incentives is to achieve the standard and no more. Under US P4P and PbR there idea is that there will be competition to provide better quality
Level (i.e. amount of reward?) – based on market price, at cost, more than cost, average cost etc	PbR/GAVI ISS reward is based on estimated average cost. QOF/CCT reward is arbitrary. GFATM based on required inputs. Some voucher schemes cover less than full cost	The higher the reward the greater the incentive to deliver. Setting the reward at average costs provides strong incentives for less efficient providers to improve. Setting the value of reward at below cost may limit incentives and compromise equity.
V different characteristics of the services being promoted	Current levels of access and cost structures differ. For immunisation coverage is already quite wide – for ARVs it is more limited. Unit costs of some services rise as you address harder to reach groups. For other services there may be economies of scale	Results are likely to be more pro poor when the better off already have access. Incentives might need to be higher for hard to cover herd to reach groups.
Degree of reliance on trust – external validation	GAVI and the GFATM use external validation. GBS uses national systems. QOF adopts a high trust model with only limited audit	Independent validation requires the use of parallel systems and additional transactions costs but should lead to more valid results and payment according to real results rather than reported results
Role it plays in income –	In Rwanda PBF covers operating costs/In Tanzania just additional costs. PbR accounts for the majority of provider revenue. QoF a large but smaller share. In P4P the share is quite small	Different implications for risks. Risks are higher the greater the dependence the scheme plays as a funding source. The higher the risk the higher the incentive the agent may need. Where share is low agent might not be interested
Who gets the money – institution or individual	For vouchers the funds go to the accredited provider (unless the	

	individual sells the voucher). Under CCTs the funds end up with the individual	
Use of sanctions for poor performance	Mostly not. Under some US P4P schemes reimbursement payment is reduce for poor performers	Potential for sanctions could sharpen incentives

Annex 7: Findings on indicators used and effects of their application

Criteria for the suitability of indicators and their application

The suitability of selected indicators to act as a proxy for performance is a complex issue. The first hurdle is whether the indicator works sufficiently well technically as a proxy of the aspect of interest.

Beyond that, the difficulties of selecting indicators and associated targets to measure progress on social aspects in the public sector have long been debated. Carr-Hill, Hopkins, Riddell and Lintott (1999) list seven potential pitfalls, quoting Smith (1993):

- 'Tunnel vision: Concentration on areas included in the outcome-related performance indicator scheme to the exclusion of other important areas.
- Suboptimization: The pursuit by managers of their own narrow objectives, at the expense of strategic co-ordination.
- Myopia: Concentration on short term issues to the exclusion of long-term criteria, which may only show up in outcome-related performance indicators in many years' time.
- Convergence: An emphasis on not being exposed as an outlier on any outcome-related performance indicator, rather than a desire to be outstanding.
- Ossification: A disinclination to experiment with new and innovative methods.
- Gaming: Altering behaviour so as to obtain strategic advantage.
- Misrepresentation: Including 'creative' accounting and fraud.' (Section 1.3.2, pp 21-22)

Some of the literature attaches the definition of misrepresentation to the term 'gaming'. However, it is useful to retain the distinction, and this paper adopts these definitions and terms in the following discussion.

Experience in the intervening years has developed these concerns. In particular, it is known that an indicator which has historically been good at reflecting performance may cease to do so when identified as a key performance indicator and used effectively in management and rewards (as in RBA and RBF schemes). The following quotation discusses 'Goodhart's Law' in the context of the health sector.

> *'Does meeting targets reflect progress in overall health system development?*
> The use of targets to spur performance is not new. Neither is concern over the distorting effects that this can have. Goodhart's Law was named after a Chief Economic Advisor at the Bank of England in the 1970s. He noticed the difficulties of attaching targets to certain measures of the money supply, stating that 'when a measure becomes a target, it ceases to be a good measure'. Once a social or economic indicator is made a target for the purpose of monitoring social or economic policy, it often appears to lose the information content that would qualify it to play such a role. It no longer represents what the government is attempting to measure. Performance targets for health service delivery such as immunization rates and trained female birth attendants have been chosen because they are thought to reflect the overall fitness of the system to attend to the needs of maternal and child health. The danger is that, once selected as targets for PBP, they cease to reflect this broader system goal and just become a measure of the ability of an organization either to meet this specific target, or fool the purchaser into believing that they have done so.' [Eldridge & Palmer pp 194 / 5]

In terms of the seven pitfalls, Goodhart's Law essentially concerns the last two: gaming and misrepresentation. There may be legitimate concerns over the other five aspects, but most of the discussion and action focuses on these two.

References are also made to cherry-picking (or 'cream-skimming'). RBF schemes may encourage managers to focus on cases that may easily be moved across the performance threshold at the expense of others who either already meet achieve it and those a long way from achieving it. (The UK education benchmark of the proportion of pupils achieving 5 or more A-C GCSE Grades is one example.) Cherry-picking means actual progress across the spectrum is less impressive than monitoring against the indicator suggests. It can also cause equity problems where the very poor and marginalised are strongly represented among those distant from the threshold. Cherry-picking is a form of gaming and also tunnel-vision.

Last perverse incentives are also a potential pitfall of managing by indicators. These may be defined as incentives to improve performance, as monitored by an indicator or indicators, that actually encourages a worse performance. An example would be where hospitals delay checking in emergency patients (and leave them outside in ambulances), in order to meet targets emergency waiting times (rather than speeding up their initial treatment in hospital, the original aim). In the schema above, perverse incentives are a particularly worrying form of gaming, where altering behaviour to improve the results is actually counter-productive in terms of service delivery.

Assessment of RBA / RBF indicators against these criteria

We now consider how far the indicators selected and their application in the schemes reviewed fare with respect to these pitfalls. Evidence where the shortcomings have been suspected, found, or not found, is reported. The assessment draws summarises material presented under 'indicators' in the templates for the schemes. Some of the pitfalls have not been encountered; they considered briefly after the others.

i) Technically

In their evaluation of the first five years of GAVI ISS dated September 2007, Abt Associates recommended:

> **'GAVI should consider additional and/or different measures of immunization performance in higher coverage countries – such as improving equity or coverage consistency.** GAVI's focus on the number of additional children immunized becomes less appropriate in higher coverage countries, as costs of increasing coverage are harder to justify in terms of disease reduction, and the amount of reward funding that countries will receive becomes lower as coverage increases and it becomes harder to immunize *additional* children'.

There is a wider point here: indicators may work quite differently in different situations.

The indicator proposed in CODA for education is the increase in the *number* of primary completers (who sit a test), rather than the completion *rate*. Thus it would reward changes due to increasing underlying population, etc., as well as any the results of success in improving education service delivery.

Technical queries have also been raised on the appropriateness of P4P indicators to all patients, such as the elderly.

ii) Tunnel vision

In the case of PbR in the UK there were concerns that providers would compromise on quality to reduce costs to remain financially viable. However while 53% of doctors responded to a BMA 2007 survey saw this as a risk, the NAO found "no clear evidence to date that this is actually happening." In particular there were fears that providers could discharge patients prematurely resulting in patients needing later readmission on an emergency basis whilst the Audit Commission (2005) did find that readmission rates were increasing – 83 per cent of PCTs (Primary Care Trusts) showed an increase in hospital readmission rates between 2003/04 and 2006/07 – but it was not possible to attribute this to PbR.

There is also little evidence that health services falling outside the ambit of RBF schemes are being undermined. A key concern is that health workers are chasing the money and only following services where additional financial rewards are on offer. Biacabe 2009 presents evidence of this in Cambodia. However, experience from GAVI and the GFATM find little evidence that it is happening. The GAVI HSS evaluation 'confirmed that GAVI's focus on DTP3 did not negatively affect measles coverage rates. These generally mirrored DTP3 coverage rates, and our model showed no statistical difference in these two indicators'. For the GFATM it was feared that its emphasis on the three diseases might be at the expense of access by other groups to other key services. However, surveys in Burkina Faso, Haiti, Rwanda and Zambia found 'no evidence of widening gaps in MCH (Maternal and Child Health) coverage by income or education'. Steele and colleagues report that care has not changed for conditions which were not included in the QOF.

There is some evidence of the quality of provision suffering where quantity is rewarded. In contrast to GAVI ISS and PbR, the evaluation of the GFATM found:

> 'basing the GFATM's PBF system largely on numeric output targets created unintended negative consequences, especially in terms of the quality of service provision. Implementers in more than half the SA2 countries reported that, on at least one occasion, they had sacrificed quality of implementation in order to achieve a quantitative numerical PBF output target.'

Implicitly the CODA Education proposal promotes the primary school sector, as opposed to pre-primary, secondary, tertiary or vocational education (which host governments may favour). It also rewards increasing access and retention rather than quality, although some would argue that a degree of quality is required to retain children in school throughout the primary phase.

A multi-country evaluation of 'performance based financing' by the Royal Tropical Institute[61] notes that the scope of the indicators has proved to be a limitation when assessing health services – typically being limited to the important programs for maternal and child health or HIV/AIDS. They regard a 'broader scope (e.g. disease control, promotional activities)' as important to ensure that the outputs come closer to representation of the desired outcome, while 'adaptation to national or local priorities, instead of global or donor priorities, are needed'. This, of course, then runs into issues of measurement and verification and has the potential to turn what would otherwise be a relatively straightforward contractual issue (which is one of the attractive features of the RBF and hybrid RBA/RBF approaches) into something quite complex.

A paper by Eldridge and Palmer notes that performance targets that are difficult to quantify may be neglected because they detract from the ability in the short term to

[61] Learning Lessons on Implementing Performance Based Financing, from a Multi-Country Evaluation Kit (Royal Tropical Institute) in collaboration with Cordaid and WHO.

deliver on specific targets. They give as an example community participation or coordination with a ministry (and/or, presumably, with a provider) not being factored into an assessment of performance because it is difficult to measure, even though it may be important for the long-term development of a sustainable health delivery strategy [ie also myopia].

iii) Myopia
While GFATM is lauded at the global level for the "transparency and focus on managing for results", the Fund's operation is less popular at the country level, with many finding the system "burdensome, rigid, and fixed exclusively on short-term outputs rather than on longer-term outcomes, results, and capacity building".

(Some other examples of myopia are mentioned in other sections.)

iv) Gaming
NAO 2008 reports:

> 'Threshold measures – measures concerned with achieving an absolute level of performance, such as the number of pupils achieving GCSE grade C or higher – are seen as particularly vulnerable to gaming. This may result in Agents concentrating their effort on those organisations or individuals who are performing just below the threshold, to the detriment of the very good or very poor performers. In some cases this may be addressed by the natural inclination of workers in the public sector to help the most disadvantaged'.

Some of the examples of other pitfalls may also be regarded as gaming. That aside, however, study of the schemes has generally not thrown up evidence of gaming, in contrast to misrepresentation.

v) Misrepresentation
Misrepresentation is a concern for many of the schemes – some more than others – and they generally take action to counter the threat. In spite of this action, there is some evidence for misrepresentation.

There is little evidence of widespread abuse in the UK under QOF[62] though this may be due, in part, to the greater risk of getting caught in a system with better oversight and might not apply elsewhere.

Olsen reports that

> 'even in Norway (number 11 in the Transparency International corruption perceptions index) there have been cases identified of systematic data fraud in the DRG coding, in order to increase resources to hospitals'

In their 2008 paper, Crouch and Mitchell consider these issues for the CODA Education proposal. It identifies the following opportunities for misrepresentation (which they call gaming):

- Allow adults to take exam
- Allow older children to take exam

[62] The median exception reporting rate was 6% in the first year of the contract, and 5.3% in the second year. One practice exception reported 86% of its patients in the first year, but this top figure for exception reporting has come down to 28% in the second year. Primary care trusts obviously have an inspection role for practices with high rates of exception reporting, but generally, there is little evidence of widespread abuse.

- Have younger children take exam
- Have out of school children take exam
- Double count children
- Import children from other regions or schools'

They note that a nation-wide system of unique student identifiers could be used to audit test-takers. However, such systems do not exist in developing countries (with the possible exception of Malawi, which is a potential pilot country). Otherwise the authors say a retest could be done in a random selection of schools. However, this would be expensive and there could be a range of practical problems with using the results. They sound the following warning note on the principles and practice proposed:

> 'one has to question the value of primary completion as an outcome indicator when the stakes are this high. Certifying completion numbers and adherence to definition is not a simple task, and, unless the definition includes an achievement requirement, it at most certifies a child's physical presence at a school over a period of time and nothing more [ie tunnel vision]. The assumptions regarding the value of primary participation alone are increasingly being challenged [technical appropriateness] '

vi) Cherry-picking

Cherry-picking is generally a concern with more complex RBF approaches in the heath sector as they create incentives to push expensive patients outside of the system.[63]

On the CODA Education proposal, the marginal costs of retaining pupils who drop out at the later stages of primary would be much lower than those who drop out in the early grades, or never enrol at all. It would also have a much quicker pay-out, perhaps by five years, depending on the structure of the education system. Management action is therefore likely to focus on near-misses, rather than those a long way away from completing primary education, which are likely to be drawn disproportionately from the very poor and marginalised [inequity].

vii) Perverse incentives

In the US the administrators of a job training programme were judged on the employment status of participants soon after they completed the scheme. This led to a situation whereby the scheme administrators were incentivised to discourage participants from seeking further training, regardless of the benefit to the participants' long-term goals, and to move straight into employment (Heckman et al, 2002, cited in NAO 2008).

A report (Journal of Health Affairs) indicates that report cards may be pushing Massachusetts cardiologists to deny lifesaving procedures on very sick heart patients out of fear of receiving a low grade if the outcome is poor.

There are concerns hospitals in the UK are being encouraged to offer to deliver services which might be delivered better and more cost effectively by other parts of the system not subject to PbR, eg at primary care.

In developing country schemes, targets for 'birth deliveries within local health facilities' may provide incentives to persuade a woman already in labour to take a precarious and potentially ill-advised journey at an inappropriate moment.

[63] Researchers found the use of such exemptions to be the strongest predictor of whether a physician. reached the performance targets and concluded, "[m]ore research is needed to determine whether these practices are excluding patients for sound clinical reasons or in order to increase income."

vii) Sub-optimization, convergence and ossification

There is no evidence for sub-optimization, convergence or ossification. Indeed, the proponents of some of the narrower schemes that focus on a single output indicator might argue that they actively free countries and service providers from constraints of conventional thinking that reflects the full results chain.

Summary of evidence of RBA/RBF indicators and usage against criteria

A mixed picture emerges. Mis-representation and tunnel vision emerge as the most common pitfalls encountered by PBA and PBF schemes. There is also evidence of perverse incentives and other gaming, cherry-picking, myopia and some technical shortcomings. In some longer running schemes, however, concerns have not always materialised as feared. The indicators for some RBA/RBF schemes appear to have been successful in avoiding these pitfalls.

In general terms, these issues are more problematic the fewer indicators are selected and the greater the stakes attached to achieving them. In particular it is hard for managers to alter their behaviour – or cheat – on a whole suite of indicators, and they are less likely to attempt to do so if the stakes are lower. Thus broader schemes such as PRBS, MDG Contracts and MCA are less susceptible to these pitfalls than focused schemes such as GAVI ISS and CODA. This is not to say broader schemes are immune: substantial funding may be strongly influenced by trends on a handful of key indicators, e.g. the incidence of poverty, especially where the payment of a performance tranche depends on them.

Perhaps better planning might have identified such adverse possibilities beforehand and allowed the design to be modified accordingly. However, it is to be expected that RBA/RBF approaches will have unintended consequences, however carefully they are designed, especially when they are implemented alongside other reforms. The key issue is to identify the problems early and act accordingly. Strong monitoring and evaluation systems will play a key role in this.

Annex 8: Household data sources in top 20 recipients of DFID bilateral aid (2008-09)

		2008/09 £ m	MICS I (95)	MICS II (00)	MICS III (05-06)	DHS Year	DHS Year	DHS Year	LSMS Year	LSMS Year	IHS ES / BS / PS Year	Last Population Census Year
1	India	297	Y	-	-	99	06	-	-	-	06	01
2	Ethiopia	166	Y	-	-	00	05	-	-	-	05	07
3	Afghanistan	147	Y	Y	-	-	-	-	-	-		79
4	Bangladesh	133	Y	-	06	99	01	04	-	-	05	01
5	Tanzania	133	Y	-	-	99	03	04	-	-	01	02
6	Pakistan	120	Y	-	-	07	-	-	05	-		98
7	Nigeria	110	Y	Y	07	99	03	-	-	-	03	06
8	Sudan	106	Y	Y	06	-	-	-	-	-		93
9	Kenya	103	-	Y	Y	98	03	04	-	-	05	99
10	Ghana	99	Y	-	06	98	03	-	98	05		00
11	Congo (Dem Rep)	94	Y	Y	-	07	-	-	-	-		84
12	Malawi	77	Y	-	06	00	04	-	04	-		08
13	Uganda	71	-	-	-	00	06	-	-	-	05	02
14	Rwanda	70	-	Y	-	00	05	07	-	-	99	02
15	Mozambique	65	Y	-	03	03	-	-	-	-	03	07
16	Nepal	58	Y	-	-	01	06	-	04	-		01
17	Burma	57	Y	00	-	-	-	-	-	-		83
18	Zimbabwe	56	-	-	-	99	06	-	-	-		02
19	Vietnam	55	-	Y	06	02	-	-	02	04	06	99
20	Sierra Leone	48	Y	Y	05	08	-	-	-	-	03	04
	Total		**15**	**9**	**9**	**17**	**12**	**4**	**5**	**2**		

Sources:
Aid statistics: Table 9, Statistics on International Development 2009, available at:
http://www.dfid.gov.uk/About-DFID/Finance-and-performance/Aid-Statistics/Statistics-on-International-Development-2009/Tables-index/

Household survey information: Annex G, Carr-Hill 2009 (Working Paper 3a for Evaluation of Fast-Track Initiative by Cambridge Education)
Updated, and augmented with Population Census column from WDI 2009, 'back matter' section, available at:
http://siteresources.worldbank.org/DATASTATISTICS/Resources/doc.pdf

Information downloaded on 24 Feb 2010

Annex 9: Sampling errors: national sample, Ghana 2008

Variable	(R)	(SE)	Number of cases Un-Weighted (N)	Weighted (WN)	(DEFT)	(SE/R)	R-2SE	R+2SE
Total fertility rate (past 3 years)	4.0	0.13	na	13,787	1.40	0.031	3.8	4.3
Infant mortality rate (past 5 years)	50.3	4.51	3,009	2,919	1.05	0.090	41.3	59.3
Child mortality rate (past 5 years)	31.1	3.81	3,037	2,950	1.11	0.122	23.5	38.7
Under-five mortality rate (past 5 years)	72.9	5.48	3,046	2,956	1.10	0.075	62.0	83.9

Appendix B, p345, Ghana DHS Report, 2005, available from DHS Macro International Website
[Extract: indicators selected by authors]

Annex 10: Evaluations of budget support

The **Joint Evaluation of General Budget Support** (JEGBS) carried out in seven countries and finalised in 2006[64] – found that:

- PGBS had a strong effect in promoting donor harmonisation, and in aligning donor programmes behind government policies spelled out in national PRSs. Moreover, these harmonisation and alignment effects were frequently found to extend beyond PGBS itself.
- PGBS has been 'unique among aid modalities' in providing holistic support to PRSs. PGBS has not imposed new policies, but has provided a forum for dialogue on how policy is implemented.
- PGBS has supported policy coherence through creating formal linkages between government policies and apparatus for linking resources to policies (e.g. Medium Term Expenditure Frameworks).
- Many of the expected effects of PGBS depend on an increase in discretionary funds available to the government budget. Whilst in the countries examined this occurred to varying degrees, even where PGBS did not clearly increase total resources available, it did lead to an increase in the volume of discretionary resources in the government budget.
- Even where PGBS was well established, the up-front transaction costs were not perceived to have fallen as much as 'some had expected'. (The UK NAO also notes that its own survey of DFID country teams suggests that more staff would be required in the short term at least to manage budget support programmes effectively.)
- Notwithstanding this, governments' transaction costs at the implementation stage were significantly reduced, by virtue of being able to follow standard government procedures rather than a multiplicity of donor ones.
- Sector ministries were encouraged to engage more seriously with the budget process (e.g. in Uganda). *The effect was weakened, however, where parallel funding to sector ministries remained significant (e.g. in Mozambique, Burkina Faso and Rwanda*).
- No obvious "crowding out" effects were found
- There was no evidence of a reduction in domestic revenue-raising effort in response to the provision of PGBS.
- Corruption was found to be a serious problem in all the study countries, but the country study teams found no clear evidence that budget support funds were, in practice, more affected by corruption, or by other fiduciary risks, than other forms of aid.
- There were some clear links between PGBS and expansion of basic services, through the additional funding available and through a collective commitment of donors and government to service delivery targets.
- PGBS had usually strengthened financial management systems.

The JEGBS also found, however, that short-term unpredictability of PGBS had been a 'frequent problem' (though this was apparently improving) and that there had been 'less progress in ensuring the medium-term predictability of PGBS (and other aid) in line with the Rome Declaration'. It also found that the scale of the benefits from PGBS was diminished by the persistence of project aid and sector baskets implemented using parallel systems to those of the government.

The **UK National Audit Office**[65] reported that:

[64] Burkina Faso, Malawi, Mozambique, Nicaragua, Rwanda, Uganda, Vietnam. A synthesis Report was also prepared: '*Evaluation of General Budget Support: Synthesis Report*', IDD and Associates, May 2006.

- General budget support has been responsible for increased "pro-poor expenditures" 'in six out of nine countries' (with insufficient evidence in the other countries).
- Budget support has increased the quantity of service delivery in seven out of eight countries, usually in basic education or health, but that maintaining quality 'has proved challenging' – and expansion in basic services has often been accompanied by a deterioration in quality (e.g. as governments seek to improve enrolment rates, pupil numbers may increase before the government has been able to recruit and train more teachers – as happened in Rwanda).
- Where macroeconomic stability existed beforehand, there is evidence that budget support has helped to reinforce it.
- In all programme submissions consulted budget support goals were specified and included measurement annexes, although 'the detail given on objectives and potential indicators varied considerably', with 15% of indicators without specific time-bound targets, and baselines absent from 22 per cent, thereby constraining assessment of progress.
- As performance based tranches form part of the design process for each budget support programme, this has led in practice to different approaches by different country teams. In India, for example, DFID budget support to the health sector in West Bengal linked around half of the funds with performance (beyond the partnership principles), compared to DFID between 2% and 10% in Mozambique.

With regard to the performance tranche, the UK NAO notes that DFID does not always have an objective or transparent way of assessing how much of a performance tranche it should disburse, which may undermine performance incentives (e.g. in West Bengal, DFID did not agree with the State Government how benchmarks would link to funding levels).

A recent **evaluation of the Multi-Donor Budget Support (MDBS) instrument in Ghana** (of which PRBS is one component) demonstrated that MDBS has had a positive influence on pro-poor policies and spending, leading to:

- an increase in the share of the education budget that goes to primary schools, for example, from 30% in 2005 to 35% in 2007;
- an increase of 22% in the number of children in public primary schools from 2.4 million in 2004/5 to 3.0 million in 2007/8;
- a rise in the primary net enrolment rate in deprived districts from 41% in 2004/5 to 73% in 2006/7; and
  an increase in the gender parity index in primary schools from 0.93 in 2004/5 to 0.96 in 2006/7.

[65] 'Department for International Development, Providing budget support to developing countries', Report by the Comptroller and Auditor General | HC 6 Session 2007-2008 | 8 February 2008.

Annex 11: Evidence of impact for major RBA/RBF vehicles

Schemes	Evidence of key positive impact/design features.	Evidence of key negative impact design features.
Budget support (PRBS and MDG Contracts) – (RBA vehicle)	Strong effect in harmonising and aligning donors behind government policies spelled out in national PRSs. Holistic support to PRSs. Practical support to policy coherence and budget-policy links. Expansion of discretionary budget resources for poverty strategies. Significant reduction in governments' transaction costs at implementation. Stronger sector ministries engagement in budget process. Increase in pro-poor expenditure. Increased in *quantity* of service delivery, usually in basic education or health. Reinforced existing macro stability. Strengthened financial management systems. No reduction in domestic revenue-raising. No obvious "crowding out" effects.	Up-front transaction costs not perceived to have fallen as much as expected (and may have increased). Expansion in basic services sometimes accompanied by deterioration in quality (while capacities catch up). Short-term unpredictability and 'less progress' in medium-term predictability (in period to 2004 at least – address to some extent through PRBS and MDG Contracts). Weak transparency in some cases over amount of performance tranche to disburse.
Output Based Aid[66] (including GPOBA) – (RBA vehicle delivered through RBF scheme)	Review showed 85% of projects achieved desired results within or below budget (compared to 49% of traditional projects).[67] Some (limited) evidence for mobilisation of additional capital. Some evidence of efficiency gains when procurement of implementing agent is competitive.	In practice, monitoring framework rarely used for purposes other than payment of subsidies. Transaction costs in practice unknown, though some very limited evidence for health PBF administration costs of between 15-30% of per capita health spend
Cash on Delivery Aid – (RBA vehicle)	(CODA remains at a conceptual/design stage. Yet to be piloted. Therefore no evidence of impact	Focusing the attention on certain sectors and disregarding others, designed to reward only good performance – what to do with non-performing countries where lots of poor people live, which might be in dire need of additional resources. Taking external factors into account can only be effective if there is at least a proportional reduction in other (*ineffective*) aid modalities assumes incentive effect will work regardless of country circumstances (de Renzio and Woods)
Voucher and	vouchers offer a great potential for	often delivered outside Government

[66] (Sparse data base of evidence for OBA. Lots of discursive information on schemes, including beneficiaries reached and expenditure, but very little information on the difference that schemes have made and their impact more generally).

[67] This may simply reflect a more realistic approach to target setting though.

Conditional Cash Transfers (RBF scheme)	donors to target the poor directly. well suited for some services e.g. maternal flexible – can cover a range of costs – direct, indirect, opportunity cost a practical transitional measure – a way of earmarking in the short term. can play a significant role in strengthening and expanding the private sector especially where provided on a large scale (as in the case of Columbia). a useful way of introducing quality assurance schemes and may lead to the establishment of a "quality culture" which could be applied within the public sector. opening the public sector up to direct competition may be a more effective way of improving performance than traditional measures such as capacity building, performance management etc	system with minimal engagement from Government – raises concerns about sustainability complex management systems – likely to be established outside Government systems can be costly (if incentive high) – can be ineffective (if incentive too low needs complementary supply side actions assumes demand side is the key constraint problems of taking schemes to scale potential to skew system to delivery of favoured services and/or overwhelm health services
UK QOF and US P4P (RBF schemes)	focus on a neglected issue – quality rather than quantity strong incentives for greater efficiency including strict sanctions for poor performance (especially in US) – risk transferred to providers rewards are proportionate to level of performance (not all or nothing) (in UK)	particular challenges in measuring quality scope for gaming e.g. abusing exempted patients option uncertainty of supply response – can pt pressure on budgets setting rewards at the right level high transactions costs – in relation to reward
GFATM (RBF scheme)	broad view of performance – responsive to country situation – use of country based indicators, consideration of contextual factors strong incentive to achieve a good performance rank risk transferred to principal recipient of funds	funds are earmarked for particular purposes (already well resourced areas) use of parallel systems – especially third party verification of results strengthening only parts of system reporting systems burdensome to capacity constrained countries – yet often not fit for purpose lack of alignment with national planning and budgetary processes GFATM does not participate in sector reviews or make commitments as part of MTEF processes requirement for additionality undermines national public financial management lack predictability in approval of progress and granting of performance payments (based on a range of criteria)
GAVI ISS (RBA scheme)	some use of national systems – initial funding provided to allow countries to make investments needed to generate results predictability in level of funding – based on self reported estimates of output – clear rules for reward payments simple and well understood payment	limited use of national systems – funding is often on plan and even on budget but GAVI does not participate in sector reviews or make commitments as part of MTEF processes funds are earmarked for particular purposes (immunisation – albeit a

	mechanism risk transferred to Government – (limited) incentive to perform reward payments are unearmarked	potentially very cost effective intervention) common global standards mean poorer countries might struggle to benefit potential for gaming – initial investment can be manipulated by self reporting
UK PbR (RBF scheme)	country owned/use of national systems strong incentive for efficiency (reducing costs) transitional measures put in place to secure support (protect possible losers); modifications made to create a “level playing field)	cannot be applied to all services need to control for gaming and unintended consequences e.g. reduction in quality, skewing on efforts towards incentivised services cost containment problems if demand response cannot be accurately gauged potential conflict with other sector objectives (e.g. for more care to be delivered at lower levels) controlling transactions cost

Group Disclaimer

This document has been prepared for the titled project or named part thereof and should not be relied upon or used for any other project without an independent check being carried out as to its suitability and prior written authority of HLSP being obtained. HLSP accepts no responsibility or liability for the consequences of this document being used for a purpose other than the purposes for which it was commissioned. Any person using or relying on the document for such other purpose agrees, and will by such use or reliance be taken to confirm his agreement, to indemnify HLSP for all loss or damage resulting there from. HLSP accepts no responsibility or liability for this document to any party other than the person by whom it was commissioned.