

# The Conceptual Basis of Payment by Results\*

Paul Clist & Arjan Verschoor  
School of International Development  
University of East Anglia

January 2014

## Executive Summary

Two economic models are used to examine the costs and benefits of Payment by Results, relative to other forms of aid. First, the principal agent model provides a framework to consider the most important factors, with a focus on linking the agent's payoff to an outcome which the principal is concerned with. Much of the promised benefit of Payment by Results is related to the efficiency improvement that comes by linking an agent's payment to an outcome of the principal's interest. Second, the multitask model draws attention to the requisite characteristics of a good measure: it is not enough to be correlated with the desired outcome. The actions needed to improve a performance measure should be similar to those needed to improve the actual outcome that motivates the principal.

The two models, and various second order effects, are summarised by a single question: do the benefits of the performance based contract outweigh the costs, relative to other forms of aid? Six headings are used to group factors which will affect the likely costs and benefits. This provides a viable framework to consider the appropriateness of a results-based contract in any given setting, underpinned by the relevant conceptual and theoretical research. Several examples are given of how the framework could be implemented, and three main research gaps are identified.

---

\*This material has been funded by UK aid from the UK Government, however the views expressed do not necessarily reflect the UK Government's official policies. Thanks are due to various DFID staff and UEA colleagues for their comments on a previous version.

# Contents

<b>1</b>	<b>Principal Agent Model</b>	<b>4</b>
1.1	Benefits . . . . .	5
1.1.1	Alignment . . . . .	5
1.1.2	Observability of Effort . . . . .	6
1.1.3	Control . . . . .	7
1.1.4	Loss aversion and naive agents . . . . .	7
1.2	Costs . . . . .	8
1.2.1	Risk Preferences . . . . .	8
1.2.2	Time Preferences . . . . .	8
1.2.3	Ambiguity aversion . . . . .	9
1.2.4	Cost of Verification . . . . .	9
1.2.5	Risk of Non-Performance . . . . .	10
<b>2</b>	<b>Multitask Model</b>	<b>10</b>
2.1	Alignment . . . . .	11
2.2	Scale . . . . .	13
2.3	Poor Performance Measures . . . . .	13
2.4	Gaming and Distorted Incentives . . . . .	13
<b>3</b>	<b>Second Order Effects</b>	<b>14</b>
3.1	Intrinsic Motivation and the Costs of Control . . . . .	14
3.2	Longer-Term Effects . . . . .	16
3.3	(Adverse) Selection . . . . .	16
3.3.1	Hidden knowledge of the counterfactual . . . . .	17
3.3.2	Effects on Agent-type . . . . .	17
3.3.3	Variability of performance . . . . .	18
<b>4</b>	<b>Principal Characteristics</b>	<b>18</b>
4.1	The Ability to Withhold Aid . . . . .	18
4.2	Time Horizon . . . . .	20
<b>5</b>	<b>Criteria</b>	<b>20</b>
5.1	Proposed Criterion . . . . .	21
<b>6</b>	<b>Worked Examples</b>	<b>23</b>
6.1	Example: Output . . . . .	23
6.2	Example: Outcome . . . . .	24
6.3	Example: Impact . . . . .	25
<b>7</b>	<b>Conclusion</b>	<b>27</b>

## Glossary

**CGD** The Centre for Global Development, a leading think tank on the issue

**COD** Cash On Delivery Aid, the CGD-term for RBA

**DFID** The Department for International Development

**DIB** Development Impact Bond

**NGO** Non-Governmental Organisation

**NHS** National Health Service

**PbR** Payment by Results

**RBA** Results Based Aid, which is a results based contract where the agent is the recipient government

**RBF** Results Based Financing, which is a results based contract where the agent is an NGO or service provider

## Technical Terms

We have carefully selected two economic models to present: both are worth the small investment as they reap rewards for thinking about the issues. A minimal amount of mathematics is used to aid clarity. Examples are used extensively in order to lower the effort required to understand the models and the technical details are omitted.

In both models, we follow convention by referring to the principal as her and the agent as him.

If reading the PDF version of this document, all of the section references are ‘clickable’ and are included to aid easy navigation of the document.

## Introduction

This document lays out the conceptual basis for Payment by Results in International Development, drawing on a large amount of academic and gray literature. Defining PbR, and related or constituent concepts, is widely noted to be difficult. Different organisations use many different terms, with boundaries between these terms suggesting a variety of different groupings. DFID have a clear use of language here, as they divide PbR into three groups (RBA, RBF and DIB) according to who the risk is transferred to. In RBA the risk is transferred to a government, in RBF it is transferred to a service provider and with DIBs the risk is transferred to an investor. This is a clear and useful distinction but it is not guaranteed that each of the three groupings *always* fit within PbR itself. As such the conceptual basis presented in this document is not necessarily applicable to the entire domain of what is currently referred to as PbR.

There are two essential elements to the outcome contingent contracts studied here, notwithstanding differences in details and circumstance.

1. There is a risk transfer as payment depends on an outcome, not action.
2. Payment is contingent on (independently) verified results.

The conceptual basis presented here is relevant in cases where both of these elements are met. First, where there is no delivery risk the interaction may be best described as simply an audited contract. This will be the case if payment is simply on an observable action. The technical distinction between an outcome and action is that actions are not fully measurable. This means that the principal pays out based on knowledge of the outcome without fully observing the constituent actions that led to the outcome. Second, where payment is not contingent on verified results, it resembles a more standard contract.

In this document we abstract from DIBS, focusing instead on RBA and RBF, for two main reasons. First, it is not clear to whom the risk is transferred so that there may or may not be an alignment between costs and income for different parties. This means that it is not clear that all DIBS will meet this basic definition of PbR (while undoubtedly many will). Second, because DIBS are newer than RBA and RBF, the evidence base is not yet established (see Centre for Global Development, 2013, for more).

The remaining document is laid out as follows. Section 1 presents the Principal Agent model which is often used to explain the underlying theory on which PbR is based. The section details, using the framework of the model, certain factors which will influence the benefits and costs of a PbR agreement. Section 2 presents the multitask model. Features of the model are discussed, with an emphasis on the performance measure itself. The desirable characteristics of a measure are presented. Section 3 discusses a variety of concerns which are not immediately apparent when using only static analysis. These *second order* effects are responses of the system to the introduction of the PbR. Section 4 summarises the relevant insights from more macroeconomic literature such as the discussion around aid conditionality. The purpose of this section is to gather relevant insights regarding the *principal's* characteristics. Section 5 synthesises many of the preceding lessons, distilling them into a criterion with six main features. Worked examples of how this framework is to be applied are given in section 6. Section 7 concludes.

## 1 Principal Agent Model

Where there has been an attempt to place PbR on a sure conceptual footing, the principal-agent model has featured extensively (e.g. Vivid Economics, 2013; Savedoff, 2010). Because of its pervas-

iveness, we sketch a simple version of the model here.<sup>1</sup> Consider a *principal* and an *agent*, where the former would like the latter to undertake some costly activity. The model was originally used to explain the interaction between a farm owner and a sharecropping farmer where the costly action was farming effort. It has since had numerous applications including between an aid donor and recipient (e.g. Azam and Laffont, 2003), an employer and employee and many others. The agent takes an unobservable action  $a$  to produce output  $y$ . For ease, let's state the production function is linear:  $y = a + \epsilon$ , so that output is simply the sum of the agent's action and a noise term,  $\epsilon$ . The output  $y$  belongs to (or is enjoyed by) the principal<sup>2</sup>. The principal pays the agent a wage  $w$  which is contingent upon output. Again for ease, let's state the wage contract is linear:  $w = s + by$  where  $s$  is a non-contingent salary and the slope  $b$  is the bonus rate. Where  $b = 0$  there is no variation in wages depending upon outcomes, where  $b = 1$  all of the variation is passed on to the recipient and where  $0 < b < 1$  a fraction of the variability is passed on. The agent's payoff is his wage minus the disutility of action  $c(a)$ , given by  $w - c(a)$ . The principal's pay off is the output minus the wage  $y - w$ . The principal must design the wage carefully, so as to incentivise effort but not reduce her own surplus.

In a standard salary contract, the wage is made up of a flat salary so that there is no outcome-contingent payment ( $b = 0$ ). The principal-agent model examines cases where there is a mismatch in preferences such that the agent does not inherently value  $y$ . In this case the principal will use an outcome-contingent contract ( $b > 0$ ) so that the agent's preferences become more aligned with those of the principal. The agent is typically assumed to be risk averse, which leads to a key insight of the model. A higher  $b$  creates stronger incentives to undertake action  $a$ , but also (for a given average wage) imposes greater risk upon the agent. The basic trade-off for the principal when considering an output-based contract is whether the positive gains from increased efficiency outweigh the extra costs in the form of the required risk premium. The increased efficiency should be obvious - the agent is incentivised to take action  $a$  in order to achieve the bonus  $by$ . The risk premium is needed because a risk averse agent will value  $by$  less because it is not certain.<sup>3</sup> In the remainder of this section we use this basic trade off as a structure in which to examine various factors, at a conceptual level, which affect the magnitude of the costs and benefits on an output based contract relative to a set transfer.

## 1.1 Benefits

What determines the size of the benefits that come from an output based contract? We discuss four issues which will affect the benefits that are apparent from considering the principal agent model.

### 1.1.1 Alignment

The main benefit of an output based contract is that the agent is newly incentivised to undertake costly actions because his payment is contingent upon output. This benefit is only an advantage over a flat salary (or fixed aid transfer) if alignment is absent or weak. Imagine, for example, an agent whose preferences are perfectly aligned with the principal save for wages ( $w$ ) and the cost of effort ( $c(a)$ ): the agent's utility is  $y + w - c(a)$ . Much of the advantage of the output-based contract in this case is lost, as the agent would undertake action  $a$  because it values  $y$ , not just because

---

<sup>1</sup>This section is based on Gibbons (1998); a very accessible account that focuses on the multitask model discussed in section 2.

<sup>2</sup>In a PbR setting, it makes more sense to state the donor benefits from the output, rather than owns it. For example, the donor may receive utility from higher immunisation rates.

<sup>3</sup>This trade-off has been recognised within the context of Results Based Financing before, see Vivid Economics (2013, p.4) and Mumssen et al. (2010, p.107).

it values  $w$ . If the agent is risk averse (diminishing the value of  $w$  as it is not certain) then the output based contract is even less attractive as the disutility of an unlucky situation where effort is expended but output does not increase (due to bad luck modelled as  $\epsilon$ ) looms large.

On a conceptual level, the current emphasis on alignment in much of the literature on PbR is surprising; the oft-cited principal-agent model implies that an outcome-based wage contract is *more* attractive if the agent is *less* aligned. However, alignment is one of DFID’s five draft suitability criteria<sup>4</sup> and it is emphasised throughout the grey literature (Vivid Economics, 2013; Toonen et al., 2009; Oxman and Fretheim, 2008). Three main reasons are offered regarding the importance of alignment.

Alignment
The current emphasis on alignment is not justified by the theoretical literature.

First, wider concerns of good donor behaviour are discussed by DFID (2013). The principles of good donor behaviour (the ‘Paris agenda’) are admirable but it is not clear that they have a bearing on the use of Payment by Results. Issues of alignment should govern whether aid is used at all, not whether PbR is used. Second, DFID (2013) argue that more aligned countries are less likely to engage in gaming performance measures (a subject we discuss later in more detail). This does offer a realistic reason why the costs of PbR will be lower if the agent is more aligned with the principal. However, alignment also affects the *benefit* of a results-based contract, to the point where complete alignment means a results-based contract is less efficient than a simple transfer.<sup>5</sup> It is difficult to imagine some midpoint where alignment is high enough for the possibility of gaming to be reduced but low enough for the incentive to create additional effort, though it is possible with strong assumptions over non-linearities. Third, both DFID (2013) and Vivid Economics (2013) raise concerns over the long-term viability of programs if incentives are not aligned. Toonen et al. (2009, p.xi) argue “buy-in from national governments is one of the major determinants that will influence the progress and scale up of performance-based financing.” The concern is again understandable, but the case is not clear. Projects that are most viable in the long run are those that are already priorities for the agent. The more the project is a priority for the agent, the less likely it is that PbR will make a material difference to project success, as well as the more likely it is that a project will be continued in the absence of aid. There is an inherent tension between trying to ensure results are additional (i.e. would not have happened without aid) and that the projects will be continued afterwards (will happen without aid in the future).

By contrast, the Centre for Global Development (2011, p.vi) does not see alignment as a pre-requisite, stating simply that the only criteria are “a good measure of progress and a credible way to verify it.” In fact they recommend PbR for fragile states, which may have low alignment and capacity. From a conceptual position the apparent benefit of PbR is foregone in the case of complete alignment, and so the position of the Centre for Global Development is more credible. We will return to the issues of gaming in section 2.4.

### 1.1.2 Observability of Effort

One of the assumptions of the Principal Agent model is that effort is unobservable: above it is described as “an unobservable action”. This is the *hidden action* problem, as the principal must contract upon output ( $y$ ) rather than the costly action ( $a$ ) itself. This assumption is worth discussing briefly. If the principal is able to observe directly the action undertaken by the agent then she can compensate the agent for this action, and does not need to contract contingent upon the outcome. This is more efficient if the agent is risk averse, as it avoids the need to provide a risk premium.

<sup>4</sup>In the most recent versions alignment has been omitted.

<sup>5</sup>The only assumption needed is that there are verification costs in the results based contract.

However, this also stops resembling a principal-agent model, and would be better described as a standard contract. The distinction between effort, input, output and outcome is rather vague in many cases but the conceptual point is simple: if the agent's costly actions can be directly observed they can be contracted upon as in a more standard market interaction.

### 1.1.3 Control

The benefit of moving to an output based contract is influenced by the degree to which the agent affects or controls the outcome (see chapter 6, Vivid Economics, 2013; Baker, 2002). This can be seen within the principal-agent model presented earlier - if the noise term  $\epsilon$  has a larger effect on the outcome  $y$  than the action  $a$ , the agent's costs and earnings are less aligned: the more noise in this sense exists, the lesser the extent to which earnings are an indicator of costs and vice versa. At the other extreme, if there was no noise whatsoever the outcome would be a direct function of effort and so the agent would face no risk as he would always be compensated for the effort exerted. Essentially effort in this setting would be observable: outcome is in this case a perfect indicator, or signal, of effort. The key contextual factor is the **ratio of signal to noise**. If the ratio of signal to noise is high then the required transfer is lower, especially for a risk averse agent. If the ratio is low then the agent will more often expect to exert effort only for misfortune to negatively affect the outcome. In the low ratio case the principal will also reward agents who exerted no effort more often - paying for poor performance.

### 1.1.4 Loss aversion and naive agents

Koszegi (2013) brings together several interesting points in a review of recent work in behavioural contract theory - contract theory in light of behavioural economics. In section 2.1 he makes the point that a more loss averse agent needs a lower level of  $b$  (lower powered incentives). Loss aversion in behavioural economics denotes the commonly observed tendency of people to experience a larger decrease in satisfaction ("utility") when something is lost than the corresponding increase in satisfaction when something of precisely the same magnitude is gained. A more loss averse agent would thus experience downward swings in  $y$  (mediated by  $b$ ) more intensely. Essentially, such loss aversion increases the incentive power of  $b$ , which is why a lower level of it is needed. Koszegi uses this insight to explain why two level incentive contracts (i.e. either a bonus is given or not) are sufficient to incentivise effort for loss averse agents.

A second relevant topic is related to the winner's curse. The winner's curse describes a phenomenon in auctions where the highest bidder probably overvalued the item, given the wisdom of crowds. Since large groups tend to get the value of an item right on average, it follows that those who valued it more than average also overvalued it. The auction set-up reinforces this phenomenon, and the principal incurs a cost as the agent (who overvalued the contract in relation to his costs) may be unable to deliver the contract. Furthermore, the principal agent model is a contract made over future actions and so naivete (an inability to accurately predict one's own future preferences) is relevant (de la Rosa, 2011; Gervais et al., 2011). For example, an agent may be overconfident regarding his ability to undertake some action: this could be included in the model as an under projection of  $c(a)$ , the action's cost. In this case the principal would be able to agree a contract with apparently good terms, only for the agent to later realise the true situation and not undertake the action (see section 1.2.5). Conversely the agent may underestimate his own ability and turn down contracts which would be beneficial.

In cases where a competitive bidding process is allied with an output based contract the former bias is of more concern as agents who overestimate their ability will tend to win contracts they

cannot utilise. How can this risk be avoided or mitigated? Agents, be they governments or service providers, with a track record in the specific domain are more likely to be able to accurately predict their own costs and abilities. In other cases, a review may be appropriate as there is little benefit in a results-contingent contract if the agreed tariff turns out to be too low, as effort will not be successfully incentivised.

## 1.2 Costs

The cost emphasised in the Principal-Agent model is the risk premium such that the agent will accept to undertake the work. This is needed in order to compensate the agent as the risk of the undertaking has been transferred to the agent. Risk preferences affect the size of the required risk premium and through analogous reasoning we may identify several other conceptually related factors which may affect the size of the premium a principal needs to pay the agent for the latter to find a contract attractive.

### 1.2.1 Risk Preferences

A more risk averse agent will need a larger premium for a given amount of risk; the more risk averse the agent, the greater the extent by which he values uncertain transfers less than transfers that are not subject to risk. As risk aversion essentially diminishes the value of the possible transfer, the strength of the risk aversion is correlated with the required size of the transfer (for a given cost). In the case of PbR it is likely that the agent will be more risk averse than the principal, as donors such as DFID have highly diversified portfolios.

One agent characteristic which we can expect to be related to their risk preferences is the agent's size. We may expect smaller agents or organisations to be more risk averse (Miller and Singer Babiarz, 2013, p.9), as a set transfer will be a larger portion of their budget and their activities will tend to be more concentrated and volatile. If the agent has few assets he may not accept a contract that has the slightest possibility of a negative outcome for the agent (this tendency strengthens with his risk aversion). Within the context of the model, this can be thought of as an agent going out of business in the case of a negative wage i.e. if  $c(a) > w$ , which is possible in the case of bad luck (negative  $\epsilon$ ). This can be seen in the Audit Commission (2012, p.16) paper in a UK context: "Small providers, including voluntary organisations or small businesses, may not be able to carry the financial risk of failure if the amount of money at risk is very large and so may not bid to provide services." Thus the risk preferences of agents will not simply affect the cost to the principal but may also reduce the size of the pool from which an agent may be drawn (see section 3.3).

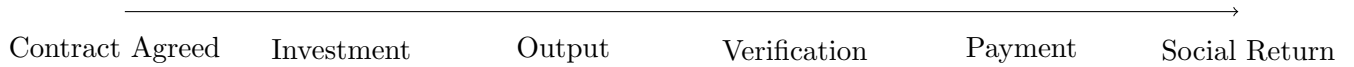
### 1.2.2 Time Preferences

Barder (2009, p.6) makes the point that Cash on Delivery aid (analogous to RBA) takes investments that will have positive social returns over the long run, and translates them into positive fiscal returns in the short run. For a graphical representation of the timeline, see Figure 1 where payment (fiscal return) precedes social return. This highlights that the investments that are targeted are often attractive investments in and of themselves, but that the time preferences of the recipient country are such that the long-term beneficial effects are not sufficiently attractive because of the short-term nature of the recipient government - their discount rate for future benefits is large. The logic is attractive but once we have acknowledged that the agent has a substantial discount rate it is not obvious that COD aid solves the problem. While it seeks to foreshorten the time between investments and social returns, when considering it as an alternative to standard project or program



aid we can see that investments are still needed before benefits are realised. In the standard aid problem the recipient government receives a transfer, makes an investment and then reaps the social returns at a later date. With PbR, the timing of the investment and aid transfer are reversed. The larger the discount rate the larger the required aid transfer, other things being equal, as the discount rate negatively affects the value of any given future payment (in the same fashion as risk aversion reduces the value of an uncertain payment).

Figure 1: Timing of Costs and Benefits



Vivid Economics (2013, pp.28-29) discusses essentially the same point within the context of RBF: it highlights the discount rate and cost of capital for recipients. Within the context of energy firms the ability to raise long-term finance on capital markets is crucial, but the conceptual analysis is the same in other settings: a high discount rate makes PbR less attractive because it places payments after investments. Discount rates or analogous agent characteristics are a key determinant of the current value of future payments.

### 1.2.3 Ambiguity aversion

Another agent characteristic that influences the value of the premium the principal needs to pay is the agent’s ambiguity aversion. From a conceptual perspective, some agents dislike ambiguity such that they would rather receive a low fixed amount than an ambiguous amount that may be higher. This differs from risk aversion in that risk has a known probability.

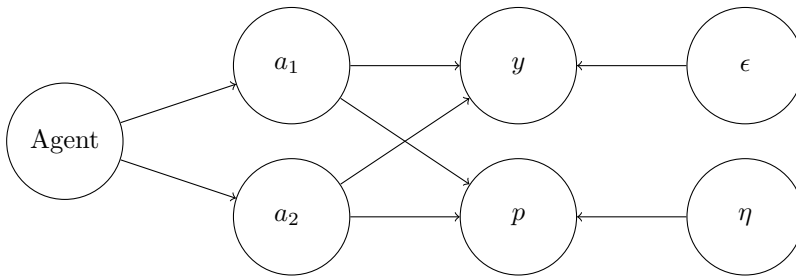
On a more applied note, it is likely that an unknown amount of future aid will be perceived of lower value if the probabilities of different outcomes are not well understood. Barder (2009, p.4) argues strongly that unpredictability is not a valid criticism of PbR: “... these concerns are... completely unfounded. Levels of school completion are much less volatile, from one year to the next, than traditional aid flows”. Barder’s defence is within the context of an RBA-type agreement based on school completion rates, and in this context the general suggested comparison seems fair - ambiguity aversion in RBA should compare volatility in traditional aid to that based on results.

While aid volatility is a valid comparison for RBA, it is not clear what the appropriate comparison is for RBF, as it is highly context dependent. For an NGO competing for funding to increase school completion rates, there are two relevant comparisons. First, the NGO will decide whether to bid for the contract or not, and so will compare an opportunity with not taking the contract. In cases where the NGO is risk averse and a set contract offers a realistic opportunity of not being reimbursed for expenditure the opportunity will simply not be pursued. Second, for the principal it is worth comparing the volatility of payments under PbR with those under more traditional forms of aid or finance for the likely bidders. The concern is that volatile income streams are used less efficiently as they cannot be incorporated into long term plans, leading to more short-term thinking.

### 1.2.4 Cost of Verification

The cost of verification is a well known problem (DFID, 2013), and will be a large factor in a simple value-for-money calculation. Some of the costs of verification are likely to have additional benefits in that they may increase the reliability of data collection systems (in anticipation of verification) but these are likely to be small relative to the cost of verification. Contexts in which verification costs are high are probably unsuitable for PbR, but this is one of the easier and more tangible costs to predict ahead of agreeing a performance based contract.

Figure 2: The Multitask Model



### 1.2.5 Risk of Non-Performance

An implicit assumption in much of the discussion above is that low output is an undesirable outcome for the principal even if she does not incur any wage costs. For example, in the case of an overconfident agent a contract may be agreed that is later found to have tariffs that do not incentivise the agent to undertake the costly action. In this case the principal has allocated funds for a given use and so must pay the opportunity costs for the public funds; this is termed the *risk of non performance* by Ghosh et al. (2012, p.16). It may be that the principal would rather have agreed a more generous contract if it had known the true nature of the agent. Non-performance may even be the case if there is no projection bias on the part of the agent, as the principal cannot guarantee that noise is absent.

Vivid Economics (2013) point out that risk transfer from principal to agent is never complete as the principal is still liable to suffer reputation risk. This covers situations that are broader than non-performance, though the principal is likely to come under scrutiny if she has agreed a contract which does not lead to the envisaged outcomes. Both the risk of non-performance and reputation risk are cases in which the principal incurs disutility even when it pays no wage. Issues surrounding non-performance are discussed further in section 4.1.

## 2 Multitask Model

Closely related to the standard principal agent model, the multitask model is highly relevant and sketched here.<sup>6</sup> The multitask model addresses a key assumption of the principal-agent model: that  $y$  is observable. Gibbons (1998, p.118) argues that the label *output* “...is misleadingly simple: in the classic model  $y$  reflects everything the principal cares about, except for wages...” Rather,  $y$  is redefined as the agent’s total contribution to firm value which cannot be measured. Here, ‘firm value’ is better understood as the true latent variable that the principal is concerned with. For example,  $y$  is not the number of students completing education, but includes the underlying aims: increased human and social capital, a more productive workforce and so on.

The assumption that  $y$  is unobservable does not preclude outcome-contingent contracts. While the principal cannot use  $y$  to contract upon, they can use the alternative proxy measure  $p$ . The wage contract is now  $w = s + bp$  and a large  $b$  creates strong incentives for the agent to increase  $p$ . A key feature of multitask models is a plurality of actions, here called  $a_1$  and  $a_2$ . Multitask models then make a variety of assumptions regarding how the two actions ( $a_1$  and  $a_2$ ) affect the performance measure  $p$  and the true outcome variable  $y$ . These relationships are shown in figure 2.

<sup>6</sup>Again, this sketch is based upon Gibbons’ (1998) accessible treatment of the model.

Imagine that both the performance measure and the true outcome are linearly and positively related to the two inputs, such that  $p = a_1 + 2a_2 + \eta$  and  $y = 2a_1 + a_2 + \epsilon$ , and the cost of the two actions is identical,  $c(a_1) = c(a_2)$ . What does this mean for an outcome contingent contract? In this case the agent's utility is  $bp - c(a_1) - c(a_2)$  where  $p = a_1 + 2a_2 + \eta$ . Note here that much as  $\epsilon$  is a noise term (luck or random variation) in the output measure,  $\eta$  is a noise term in the performance measure. The second action  $a_2$  is twice as effective as the first in increasing the performance measure with the same cost, and so the agent will allocate more effort to undertaking that action. However, the first action  $a_1$  is more beneficial for the latent variable (outcome) and so the incentive distorts effort by over-rewarding  $a_2$  and under-rewarding  $a_1$ .

### An Example Using Completion

Before moving on to Figure 3 (see Neal, 2011) it may be useful to consider an example. Imagine a performance based contract where a donor agrees to provide an incentive to a recipient government based on student completion rates, because it has the underlying goal of an educated and productive workforce. The agent has two actions: it can either increase the quality of education provided, or can strongly discourage drop out or repetition, even in cases where these may be beneficial. Intuitively it is obvious that both actions may be desirable but the balance between the two actions will be different, depending on whether the observed performance measure is aimed at or the unobserved (i.e. latent) true outcome of interest. Cracking down on repetition would increase completion but is not the same as increasing the number of educated students<sup>a</sup>.

<sup>a</sup>Note that completion in this context commonly means the numbers taking an exam, not those passing it.

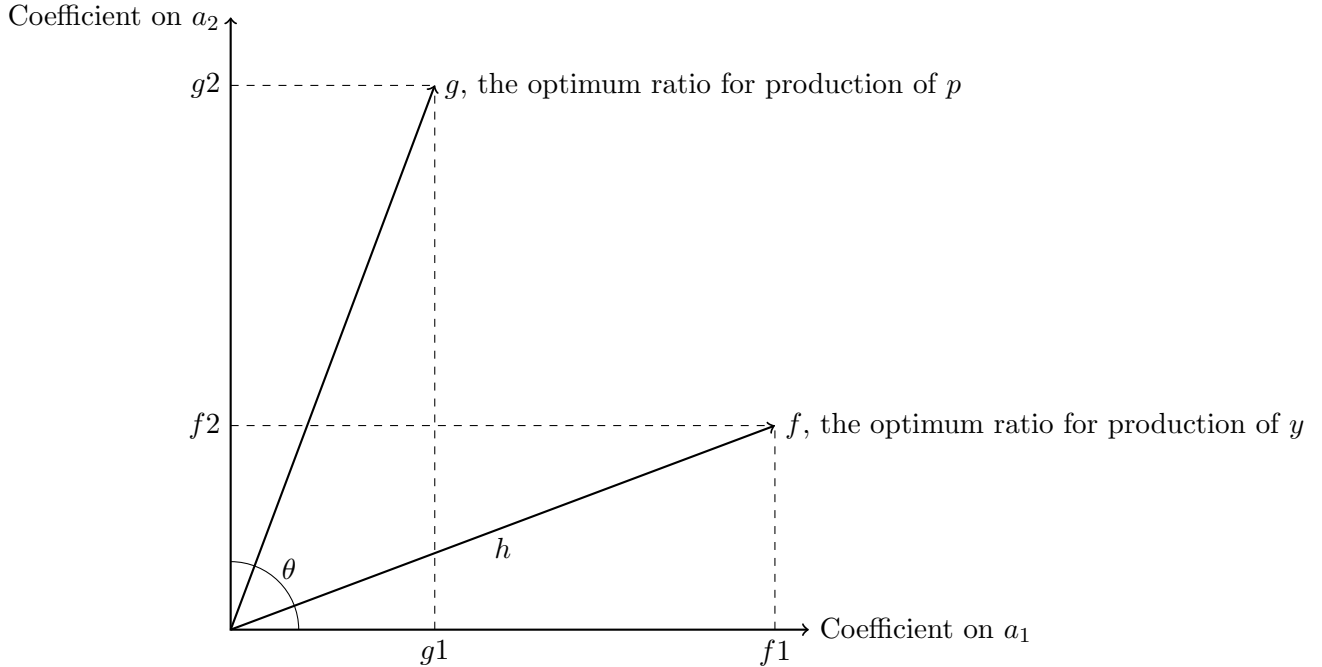
Consider the example in the box. To restate it in the language of the model,  $p$  is the completion rate,  $y$  is the underlying aim of a productive workforce,  $a_1$  is an increase in quality and  $a_2$  is a mechanistic crackdown on drop out and repetition. The first action has a small effect on the performance measure  $p$  but a large effect on the latent variable  $y$ , and vice versa for action 2. Figure 3 plots the implied coefficients in the optimum wage contract. Point  $f$  shows that the required mix of actions to achieve a productive workforce ( $y$ ) is weighted towards rewarding an increase in the quality of education ( $a_1$ ) and would only weakly reward more stringent action on repetition and drop-out ( $a_2$ ). If one wanted to design the optimum contract to achieve completion as a goal in itself, point  $g$  shows the relative mix of the two actions would be reversed.

## 2.1 Alignment

The optimum incentive contract (level of  $b$ ) rests on two factors: alignment and scale (Neal, 2011; Gibbons, 2010). Alignment here is about the latent variable's response to actions versus that of the performance measure. It is greater where angle  $\theta$ , between the lines  $g$  and  $f$ , is smaller. Technically, the optimum  $b$  (as discussed in the principal agent model) is partly determined by  $\cos\theta$ . To see what this means consider the two extremes. If the two lines are perfectly aligned (such that the angle between them is 0)  $\cos 0 = 1$ , and so the optimum  $b$  would be one i.e. if the optimum ratio of the two actions for producing the latent variable and performance measure are identical, all of the variability in outcomes can be transferred to the agent. By contrast, if the two lines are orthogonal such that the angle between them is a right angle, the optimum  $b$  is 0 as  $\cos 90 = 0$ . In other words, if there is no overlap between the required efforts for the performance measure and the latent variable, there should be no outcome-contingent contract as the performance measure is so poor.

Intuitively, one may think that the important consideration with a performance measure is that

Figure 3: A Multitask Model with Two Actions



it is highly correlated with the latent variable. As Baker (2002, p.736) puts it: “[w]hat matters is not whether the measure is correlated with firm value, but whether the margins on the measure are similar as captured by  $\cos\theta$ .” In other words, a focus on the correlation between the latent variable and performance measure may mislead one into thinking a performance measure will be suitable when it is not. It is more useful to think about the required mix of actions that lead to improvements in the two measures. To return to the completion example, the conceptual problem is that increasing quality may have a smaller effect on completion than it does on productivity. A focus simply on the normally high correlation between completion and a productive workforce may be misleading, as it does not reveal likely distortions.

The preceding argument can be recast in standard economic language, by noting simply that the correlation between a performance measure and an outcome is observed *ceteris paribus* i.e. holding other things constant. Once the performance measure is selected things are likely to change, and the previously observed correlation may no longer hold as an agent’s behaviour adapts to incentives to produce  $p$ , and the associated ratio of the two actions. Goodhart’s law (Eldridge and Palmer, 2009, p.164) sums this up neatly: ‘when a measure becomes a target, it ceases to be a good measure’. These insights have empirical support: Courty and Marschke (2008) tested the covariance of performance measures and proxies for the latent variable, and found that when a measure is incentivised it has lower covariance with the true latent variable. In simpler terms, the insight of the multitask model is that the measure is good not if it is normally correlated with the latent variable but rather if it induces valuable actions by rewarding the right mix of actions.

Gibbons (2005, p.5)

“More generally,  $p$  is a valuable performance measure if it induces valuable actions, not if it is highly correlated with  $y$ .”

## 2.2 Scale

The second point emphasised in the literature is to do with the scale of the incentives. Imagine that point  $h$  in figure 3 represents an alternative performance measure to that given by  $g$ . This second performance measure is completely aligned with the outcome  $y$ , with the preceding section implying the optimum contract has high-powered incentives ( $b = 1$ ). However, the *scale* of  $y$  and  $h$  are not the same - in order to produce the improvement in the performance measure  $h$  the agent only needs to exert half of the amount of effort as it does in producing the increase in  $f$ . In other words,  $h$  will rise and fall with  $f$ , but by only half the amount. In this case the principal would need to change the scale of the incentives so that the agent is rewarded with the increase in  $f$  if it increases the performance measure  $h$  by two.

## 2.3 Poor Performance Measures

The title of Kerr's (1975) article neatly summarises the pessimistic nature of one strand of the literature regarding poor performance measures: 'On the Folly of Rewarding A, While Hoping for B'. If the performance measure is poor, the principal does not have many options to improve its lot. The model simply states that in cases of low alignment the outcome-contingent part of the wage should be low (i.e. decrease  $b$ ) as weak incentives are more efficient than strong incentives for the wrong action (Lazear, 2000; Baker, 2002).

Neal (2011) takes a more positive line in such situations, arguing that some level of incentive pay is still beneficial as long as neither action has a detrimental effect on either variable. This has a solid conceptual basis, but rests on several assumptions including a complete lack of alignment between the preferences of the principal and the agent (we return to the case where the agent does have some intrinsic motivation in section 3.1).

## 2.4 Gaming and Distorted Incentives

The multitask model provides a framework in which to think about gaming, and distinguish it from more complicated unintended consequences. Baker (2002, p.729) discusses a good example of gaming from Fast and Berg (1975): Lincoln Electric abandoned paying a piece rate to their secretaries based on the number of key strokes when it was discovered they were using lunchtime to tap a single key repetitively. This is clearly gaming as the repetitive tapping of a key (call it  $a_2$ ) had an influence on the performance measure (the number of keystrokes, say  $p = a_1 + a_2 + \eta$ ) but not on the latent variable ( $y$ , where  $y = a_1 + \epsilon$ ). We can confidently call this gaming as in this case the  $a_2$  served *no useful function* in terms of the firm's total value. The possibility of gaming should not be taken lightly, as there is often evidence of actions which have no positive (or even a negative) effect on  $y$  (Bevan and Hood, 2006, provide several examples from the NHS). The insight from the multitask model regarding the correlation of two measures being a misleading predictor of measure success is very relevant here, as the following quote illustrates:

"The amount of bureaucratic gaming is not predicted by the correlation of measured outcomes with desired outcomes, but rather by the correlation of the effects of employee actions on measured outcomes with the effects on true objectives. Unfortunately, this latter correlation is much harder to estimate." Baker (2002, pp.736-737)

In the case of the typists, before the number of keystrokes was selected as a performance measure, it was probably a very good indicator of secretary output. However, the number of keystrokes could readily be increased in a way that had no effect on actual output, and so the measure stopped being correlated with true output.

While gaming receives a lot of attention, there is a different class of unintended effects that are less morally suspect and probably more important and widespread. These are effects which may not be obvious when considering only a static model or situation, and are often unintended and unforeseen consequences.

To illustrate distortion effects consider two examples. First, Sylvia et al. (2013) found that in China the provision of incentives for reductions in student anemia in schools may have displaced teaching effort, leading to lower test scores in some cases. Second, Barder and Birdsall (2006, p.17) state that “[a] serious risk in the case of education is that of a shift in priorities and resources towards quantity over quality.” In both of these examples, there is a genuine concern that  $y$  may be negatively affected: the quality of education in both cases is threatened. Neither increasing the quantity of school places nor fighting anemia are wrong *in and of themselves*, but the second order effects of incentivising either of these intermediate goals may negatively affect the latent variable through distorting incentives such that they receive an undue weight. This real-world example closely resembles the situation in figure 3, where the performance measure overweights one action. In this case, the introduction of the measure led to a decrease in provision of education time ( $a_1$ ) and an increase in school meals ( $a_2$ ).

#### Gaming and Distortion

Gaming receives a lot of attention because it is morally dubious. However, distortion of incentives should, in most cases, be of greater concern.

### 3 Second Order Effects

There are a number of reasons to be sceptical of the promised efficiency improvements laid out in the principal-agent model. Some of these reasons have been discussed in the context of the multitask model, and stem from the ubiquitous problem that the true latent variable which a principal is concerned with is unobservable. In this section we discuss a separate group of ideas which stem from the ceteris paribus nature of the Principal-Agent model. Ceteris paribus means that analysis is done ‘holding other things constant’ and is a useful simplifying assumption. However, there are several reasons to think that other things are not constant in this situation, but rather the system adapts to the use of a performance based contract. Following convention, we call these adaptations that become visible once the ceteris paribus assumption is relaxed “second order effects”.

#### 3.1 Intrinsic Motivation and the Costs of Control

Akerlof (1982) noticed that firms often pay wages that are higher than they absolutely need to, and that workers in response often expend more effort than needed. This observation was a challenge to the assumption of pure profit-maximising individuals. He explained this as a ‘gift exchange’ such that the worker feels an inherent obligation or desire to provide a degree of effort that corresponds to the salary. Since these early insights a large body of related literature has examined issues of effort and payment, using many different terms. Here, we will discuss strategic ambiguity, intrinsic motivation and the costs of control.

Much as Akerlof (1982) noticed that firms and employees did not follow traditional economic precepts when setting their wage or effort levels, Bernheim and Whinston (1998) noted that many contracts were less complete than expected. They argue that this is not an oversight or anomaly but rather a rational decision; termed *strategic ambiguity*. To simplify slightly, the ambiguity in contracts is useful because of two assumptions: a) performance may be observable but non-verifiable (i.e. both parties know when a clause has not been met, but there is sufficient noise that it would be difficult to enforce in a court) and b) that all elements of the relationship are strategic. The key

insights for PbR are that a transparent payment rule precludes an ability to respond to broader signals regarding the effort, need and ability of the agent (at the very least to the extent of the money committed). It may be the case that the control gained by a contract is outweighed by the loss of the ability to strategically respond to the situation.

To further develop this argument, imagine that there are many possible performance measures for a given action that are all highly correlated with the output. However, all have low alignment and would thus distort incentives if used as the sole measure. Even a set of performance measures may cause distortion in this case, as each may be subject to gaming or similar concerns. In this setting, strategic ambiguity offers several advantages. The agent does not know exactly which performance measure will be used, but does know that future contracts will be rewarded on the basis of past performance, and that the principal's true concern is  $y$ . The agent's best response to this situation is to focus on producing  $y$ , or some of the likely performance measures. In this case, there is little distortion in incentives as the agent is focused on gaining future contracts but expends little energy in affecting specific indicators. The agent's response to this situation would depend upon the principal's ability to choose wisely amongst the agents, and the regularity of contracting.

*Intrinsic motivation* can be included in the multitask model; Neal (2011) introduces the idea that there are norms dictating the amount of time or effort devoted to  $a_1$ , and that deviation from this norm is (increasingly) costly. Gertler and Vermeersch (2012) make a similar argument in their model of performance incentives in health, when they argue that there are minimum ethical levels of care for both inputs and even if only one is incentivised and the two goods are substitutes, the ethical minimum will mean that neither good will stop being produced. To give a concrete example, this model argues that if only TB vaccinations are incentivised, the ethical standards of the health practitioners would mean that treatments of malaria do not drop below a basic standard.

**Financial Incentives**

Many terms have been used to describe very similar concepts. The common insight is that financial incentives are not independent of intrinsic motivation, and the relationship between them is complex.

Neal (2011) points out that if the ethical minimum is very low, the incentive pay may well still be sensible even if  $a_1$  and  $a_2$  are substitutes. This is because the benefit from the incentive outweighs the distortion. Muralidharan and Sundararaman (2011) argue that there is evidence of very low effort provision in their setting, citing high absenteeism in the Indian education sector. The more general conceptual point is that where there is some intrinsic motivation (often called alignment in a macro setting), the distortion from financial incentives may have a detrimental effect on the true latent variable as weaker incentives for  $y$  may result in better outcomes than stronger incentives for  $p$ .

*Prima facie*, one may think that adding extrinsic (financial) incentives to intrinsic motivation would be a sensible combination. However, there is a body of evidence (reviewed in an accessible way by Gneezy et al., 2011) that states extrinsic motivation can crowd out intrinsic motivation. Gneezy and Rustichini (2000) showed that when small fines were introduced for parents in Israel who were late to pick up their children from daycare, the number of parents showing up late *increased*. This is explained by a shift from intrinsic motivation where arriving late is not fair on the employees, to a market orientation where payment of the fine entitles the parent to arrive late. In this case the extrinsic motivation was not additional to intrinsic motivation, but replaced it.<sup>7</sup>

<sup>7</sup>DFID have started to withhold a fraction of payment in cases where an output is late in being delivered. Evidence would suggest this will increase the number of outputs delivered on time, but also increase the size of the delay. This latter effect is similar to the evidence from the Israeli daycare centre - once a fine is paid a contractor may feel under no compulsion to complete quickly as the fine has already been incurred. The net effect is not clear.

Falk and Kosfeld (2006) provide laboratory evidence of what they term the *hidden costs of control*. In the experiment a principal is allowed to set a minimum performance requirement for the agent. Standard economic theory would predict this minimum would be sensible to enforce, and that it would have only beneficial effects. In cases where principals did enforce this minimum threshold, agents' performance was *lower*. This is explained as a response to a signal of distrust from the principal. In cases where the agent was allowed to choose any effort level, performance was higher. Agents in this setting appear to be *control averse*.

The degree of monitoring and the sense of fairness appear to be crucial in the literature (Dickinson and Villeval, 2008), though there is some way to go before such insights can be applied easily to the PbR setting, as most of the empirical or field work has been at the individual level, and not at the organisation level. However, the conclusion from the body of evidence is that the use of performance based contracts can have unpredictable effects on performance, and that they should be employed with care and attention.

### 3.2 Longer-Term Effects

We now move on to various issues that would be apparent over a longer time horizon, which may reveal a greater disconnect between the performance measure and the true output. Darley (1991) echoes this analysis discussing the introduction of performance measures:

“When we did this, we unleashed a complex set of social processes.... [We] caused all the players in the system to turn their attention to maximizing those numbers. Sometimes - often - they did so in ways that we did not anticipate, and that caused actions that “maximized the numbers” without maximizing what we really care about. In fact, the act of imposing the control system often destroyed what we really cared about.” Cited in Baker (2002, p.739)

These quotes can be thought of as instances in which a variety of actions are discovered which do affect  $p$  but do not affect  $y$ . As time passes effort is put into seeking out new actions which incur low costs but have large marginal benefits on  $p$ , regardless of their effect on  $y$ . In this case the distance between  $y$  and  $p$  grows as innovation and research is misaligned.

Another longer term effect relates to how the incentive is perceived over time. To give a micro example, consider the following quote:

“Health workers began to perceive this bonus as a right and expect it every month, even when performance did not justify the payment of the entire bonus, and they reportedly spent time on endless discussions to obtain it, rather than in the provision of services to the population.” Bertone and Meessen (2012, p.7)

The difficulty of these longer term effects is that they are hard to predict.

### 3.3 (Adverse) Selection

Conceptually, adverse selection is a case where asymmetric information (where different parties have different levels of information) leads to ‘bad’ outcomes in a market transaction, from the perspective of the principal. For example, if an insurance company offers a very high level of health care for a given cost it will be most attractive to individuals who have a high need of healthcare: the most expensive people to insure. In the healthcare setting the information asymmetry refers to the fact



that individuals may know more about their need for healthcare than the insurance company.<sup>8</sup> This is relevant in a PbR setting for three reasons.

### 3.3.1 Hidden knowledge of the counterfactual

It is difficult for the principal to predict counter-factual performance against a performance measure in the absence of PbR. This means that it is difficult for the principal to know whether it is paying for additional success in an area or improvements which would have happened anyway. The agent may have private knowledge over coming improvements against some measure which the principal is not aware of.

This difficulty is compounded by a trade-off in the setting of the tariff. DFID's (2013) draft suitability criteria include reference to the likely attraction of the extra incentive, and the feasibility criteria include a reference to value for money. These are in tension - a higher incentive may increase cost and the likely response to the incentive, and vice versa. The problem of adverse selection makes the setting of a tariff more complicated, given the agent is better placed to predict future trends than the principal.

### 3.3.2 Effects on Agent-type

Over time the type of agent that bids and agrees to a performance based contract may change (in an RBF setting). This is viewed very positively in some of the academic literature, with Lazear (2000) stating that much of the observed 44% increase in factory output was associated with a change in the type of employee. Inefficient employees tended to leave the firm, and the quality of workers coming into the firm increased. In a more theoretical piece of research related to performance-based pay for teachers Lazear (2003) argues that the selection effect is one of the two major drivers of improvements (alongside straightforward incentive effects). This can be seen in the context of the principal agent model. Imagine a number of agents with different costs of effort (alternatively, this could be explained using different efficiency levels). Over time agents with lower costs of effort will enter the market, and agents with high costs of effort will leave the market.

#### Effects on Agent Type

Over time, we would expect agents with *lower* levels of intrinsic motivation to prosper if performance based contracts are consistently used.

However, this result should not be extended too far, and the multitask model provides a useful framework to think about the limits of this reasoning. As Gibbons (1998, pp.117-118) points out, commenting on the Lazear (2000) article, "not all jobs are as narrowly defined and easily monitored as windshield installation." It is more plausible that over time the quality of agents does not increase according to the metric of their ability to produce  $y$  but according to their ability to produce  $p$ . In RBF this may mean agents which are able to increase measured performance are attracted into a sector or profession, and the link between  $y$  and  $p$  may be influenced by the type of agent.

Within the multitask framework where competing agents have different degrees of intrinsic motivation, we would expect agents with *lower* levels of intrinsic motivation to offer more competitive proposals because they are less scrupulous about gaming performance metrics. In other words, over time we would expect the number of agents who are happy to game performance metrics to increase and the number of agents who share the principal's concern for the latent variable to decrease. This

---

<sup>8</sup>The classic example is from Akerlof (1970) who introduced the idea in the context of a second hand car market, where sellers know more about the true worth of the car than buyers. This is called the 'lemons problem', where buyers are worried about purchasing a very poor quality car.

is because agents with no intrinsic motivation are not distracted by actions which are less efficient in terms of increasing  $p$ . In cases where  $p$  is equal to  $y$  this concern is not relevant.

### 3.3.3 Variability of performance

There are more short term concerns regarding agent characteristics such as variability in agent performance. Imagine there are several agents bidding for a contract who differ in terms of volatility in their performance but who have the same average cost of effort. One key difference between them might be that some agents take more risks than others in the execution of contracts, which sometimes pays off and other times doesn't. It is easy to imagine a situation in which the agents with greater volatility in their efficiency would be able to produce a more attractive financial tender: let's say they gamble on their ability to deliver at low cost knowing they can choose not to invest and thus not to deliver on a contract if they discover their costs to be high. This could be suboptimal for the principal as she may prefer to agree a higher tariff with an agent that will exert effort in all circumstances.

## 4 Principal Characteristics

Much of the literature discussed in the preceding sections has used analytical tools developed with labour contracts in mind: both the principal-agent and the multitask model were originally used to analyse an employee-employer relationship. This section switches from microeconomic literature to macroeconomic research, using the aid conditionality literature which considers bilateral aid relationships<sup>9</sup>. This section is inevitably more focused on RBA-style agreements. Two key insights are discussed which deal with the characteristics of the donor. First, it is assumed in much of the microeconomic literature that the principal is able to withhold payment in the case of non performance, but the macroeconomic literature questions this assumption. Second, in light of the second order effects, the principal's ability to contract over longer time periods is discussed.

### 4.1 The Ability to Withhold Aid

The question of how aid should be allocated and used has been extensively studied. Collier and Dollar (2002) argue that it should be given in response to both low levels of income and high levels of quality of governance. There is an inherent difficulty in following this advice. The quality of governance is typically defined, at least partly, according to its ability to increase growth over the long term. This reveals a tension inherent within the advice: poor countries tend to be less well governed and so donors typically choose to allocate more aid to poorer countries, and good governance is defined relative to levels of poverty, not in absolute terms. The response to these twin allocation criteria is that aid is almost invariably allocated to countries where the quality of governance leaves much to be desired.

Ex Post conditionality was a common solution to the allocation problem - aid was granted in exchange for certain promised policy reforms. It has been well documented that this seldom worked as planned (Svensson, 2000; Collier et al., 1997). Recipients soon learned that if conditions were not met, donors did not follow through on promised 'punishment' of withholding aid. This has usefully been termed the Samaritan's Dilemma (Buchanan, 1975), as the principal's desire to help the agent means they are unable to withhold help (inequality aversion). Another reason is more institutional - Easterly (1999) argues that the incentives for individuals in aid organisations reward disbursement and make withholding aid difficult.

---

<sup>9</sup>For accessible reviews, see Temple (2010) and Morrissey (2013).

Ex Ante conditionality is a related idea: the donor chooses to allocate aid not on the basis of promised performance but on the basis on past performance. Collier and Dollar (2002) are exponents of this approach, but the weight of evidence is that governance or policy measures are insignificant in aid allocation decisions (Clist, 2011; Hout, 2007), probably because of the tension between the twin aims. The track record of (both ex post and ex ante) conditional aid shows that it is difficult for aid donors to allocate aid with reference to governance levels. Other concerns (income, population and wider links) seem to be much more important.

#### An Important Donor Characteristic

If the donor is unable to withhold aid from the recipient (even if it is within a separate modality), the power of the incentive is lost.

The above is highly relevant in the case of PbR as its underlying logic assumes that the principal is able to withhold aid in cases of non-performance. The concern from the grey literature is well expressed in the following quote:

“‘Cash on delivery aid’ is designed to reward good performance. Better performing countries will receive more aid, but this still leaves open the issue of what to do with non-performing countries where lots of poor people live, which might be in dire need of additional resources. Should they be disregarded? If so, can it be assumed that that will provide a positive incentive for the government to redouble its efforts in pursuit of better indicator scores? If not, should other aid flows be able to compensate them for the loss on ‘cash on delivery’ allocations? And if that happens, does that then not completely undermine the case for ‘cash on delivery’ aid, given that countries know that they will receive aid regardless of performance?” De Renzio and Woods (2008, pp.2-3)

The inherent tension between a desire to reward performance and alleviate desperate situations is apparent. If countries with poor performance receive aid anyway, the incentive is lost. Conversely if these countries receive no aid then they are ‘disregarded’. One of the reasons for this tension is the difficulty of separating effort from conditions. If a country’s poor performance is due to poor conditions it should receive more aid, but if it is due to low effort it should receive less. The signal/noise ratio is low, and so the task of separating between the two is difficult. The tension can also be seen in the following quote on contingencies in COD aid agreements:

“Contingencies are still necessary, however, for events beyond the recipient’s control that could interfere with schooling or disrupt testing - such as major natural disasters or declines in the world price of an important export. While countries facing such crises may require additional aid, it is important for those funds to come through other channels.” Centre for Global Development (2011, p.63)

The message is clear: in the case of factors beyond the recipient’s control other forms of aid should be used to compensate the recipient. However, the practical matter of deciding which factors are beyond the recipient’s control and how much to compensate is difficult. Does a 1% fall in the world price of their largest export warrant a payment? What if the donor believes the recipient has been unwise in not diversifying its export market: should that be taken into consideration? This is important as there are many possible confounding factors which could conceivably be claimed as mitigating circumstances.

The analysis above suggests that the institutional ability to withhold payment may be low, which may be compounded by more practical matters. To turn to the latter, DFID’s draft internal documents (e.g. DFID, 2013) raise two main concerns. First, the financial systems used require

clear annual budgets and do not cope well with overspend or underspend. Second, unutilised funds in PbR settings are typically expected to remain within the country. If the donor is unable to withhold aid from the recipient (even if it is within a separate modality), the power of the incentive is lost.

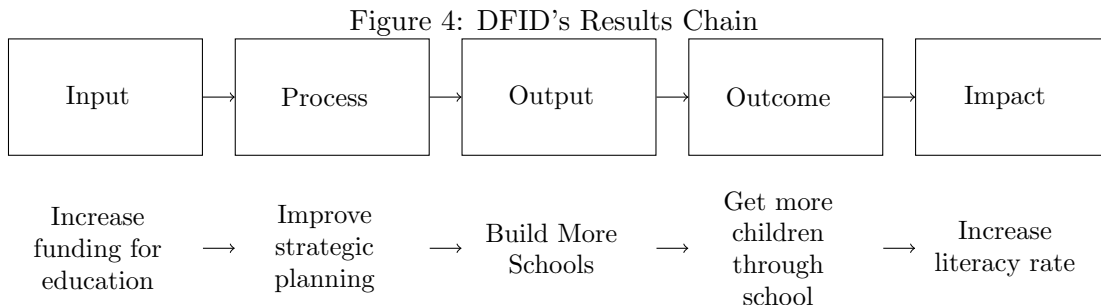
## 4.2 Time Horizon

The Centre for Global Development (2011, p.20) suggest a *minimum* contract period of five years, as shorter periods of time will impinge upon the ability of the recipient to innovate. While the logic of a longer time horizon holds, the ability of the donor to agree to it is not guaranteed. Note that the time preferences of the recipient (section 1.2.2) are not necessarily a binding constraint here as there are annual payments - a longer contract would mean the recipient could predict his future income with greater accuracy.

It is easy to see that when donors are not able to commit themselves for sufficiently long time periods, the likelihood decreases that the recipient undertakes investments that take longer to come to fruition. This concern becomes binding if we make plausible assumptions regarding the time between an investment and its return. Recall the example in section 2 where a recipient is being incentivised to increase its completion rates. The two possible actions in the stylised example are to 1) improve quality and to 2) discourage drop out and repetition. The time needed for quality improvements to lead to increases in completion is high, but the required time for a mechanistic clamp-down on drop out and repetition is relatively low. In this case a donor which agrees a short-term contract would make it more likely that the less good action would be taken.

## 5 Criteria

Before presenting the suggested criteria, it may be useful to briefly discuss a framework commonly used within DFID: the results chain. Figure 4 shows the results chain as discussed in DFID (2011, p.11) with the 5 levels in boxes, with an example of each underneath. Clearly, the results chain is a simplified conceptual framework with a specific end use in mind. The provided examples are useful in this regard, especially as larger projects (over £1 million) are expected to have several indicators at the output, outcome and impact levels. In the examples provided the three levels have obvious indicators: number of schools built, number of children completing school and the literacy rate in the general population.



There is a clear difference in emphasis between the results chain in figure 4 and the multitask model presented in section 2. The multitask model *expects* the performance measure and the latent variable to be different, and focuses attention on the vital importance of assessing the measure. The results chain does not mention this, and the selection of a good measure could be seen as a detail

or formality. The emphasis in the results chain is on the different levels, from a simple increase in funding for education through to an increase in the literacy rate. The results chain offers an attractive simplification of a theory of change, and can clearly be useful. However, there are a variety of assumptions which are implicit in the results chain, and a greater emphasis on the exact relationship between the desired outcome and the performance measure would be beneficial. The guidance to provide several measures at each level of output, outcome and impact may be naive in some contexts.

The results chain inevitably leads one to think in a linear fashion, whereas a consideration of the system as a whole may be more useful. The distinction between inputs, outputs, outcomes and impacts is often vague, as many could be rethought of as a different level in a different system. For example, ‘build more schools’ is thought of as an output in figure 4. It could conceivably be thought of as an outcome, where the output is to increase the number of skilled builders in a country. Likewise, completion is discussed above as an outcome, but could be seen as an impact in itself. Higher completion could even be thought of as an output in a broader system where a skilled labour force is the outcome and greater economic growth is the impact. These definitional issues are not irrelevant details but may have real policy significance. The multitask model provides a framework where the focus is on two variables (the latent variable and the performance measure) and the factors which affect them (be they actions under the agent’s control or noise).

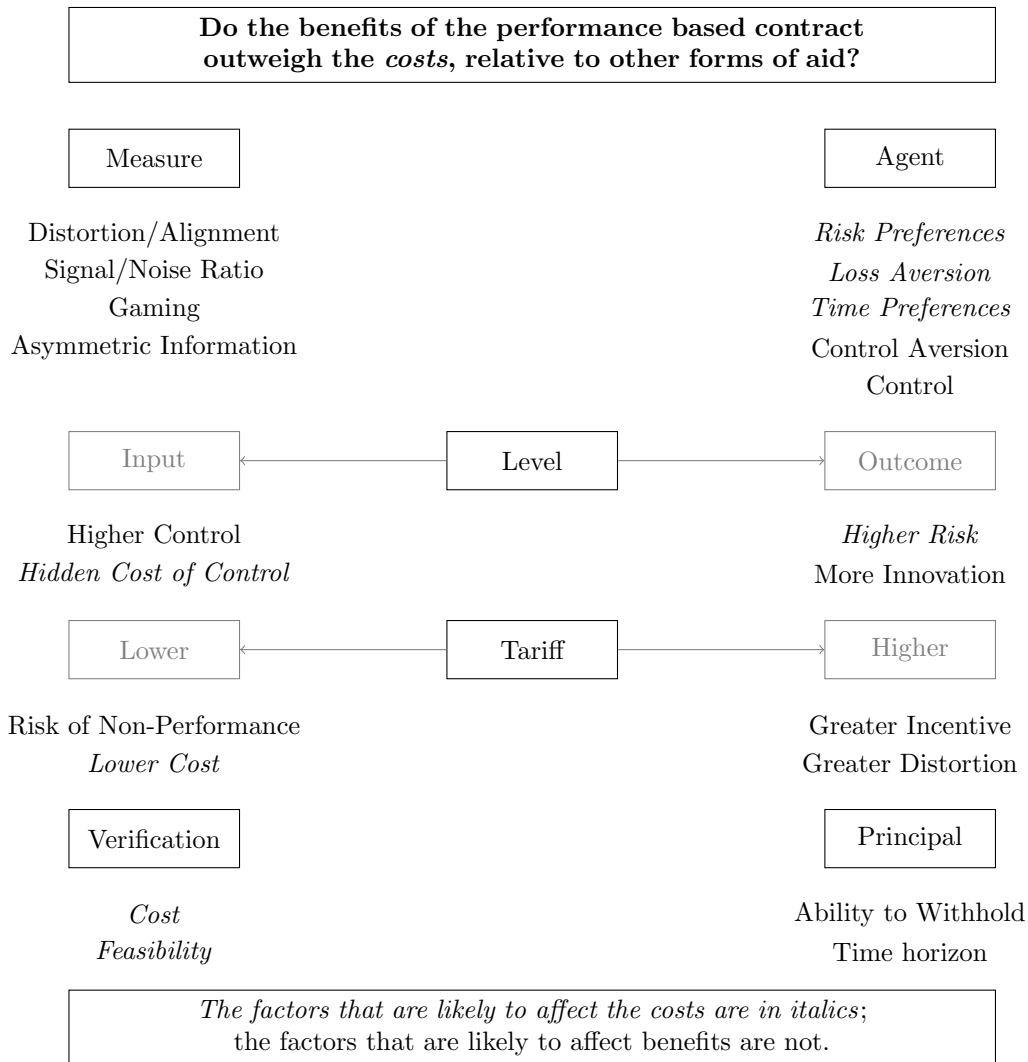
To illustrate the possible danger of the results chain, recall two examples from section 2.4. The ‘number of key strokes’ is a plausible measure for output in a secretarial business and could be part of a reasonable theory of change and results chain. The repetitive key-tapping that undermined this intervention was a good example of a gaming action: an action that increases the performance measure without changing the latent variable. While theory-of-change frameworks often regard gaming as a risk, they tend to be more focused on clarifying assumptions regarding the relationship between the specific steps than between the measure and the thing it is trying to measure. Nor do they take account of the wider effects on other activities. Recall the example of Chinese schools being incentivised to reduce anemia. The linear nature of the results chain would have been accurate in describing an improvement in the school performance of anemic children. However, non-anemic children saw their academic results drop slightly, most probably due to a reallocation of resources away from teaching time. The distorted incentives would not become apparent by using a results chain analysis as the other parts of the system are excluded from the analysis. This is also a cautionary tale to bear in mind when judging evidence of PbR success - if the only indicator of effect is the measure itself, important effects may be missed and success may be illusory.

## 5.1 Proposed Criterion

In figure 5 we suggest only one criterion: do the benefits of PbR outweigh the costs, relative to other approaches. These factors are highly context dependent but the remainder of the figure allows a policy-maker to think through six elements which will determine the likely benefits and costs relative to alternatives. The conceptual discussion of each of these factors can be found in the preceding sections; section 6 shows in outline what implementing this would look like in practice. The promised benefit of PbR comes from transferring the risk of delivery to the agent, thus aligning his benefits with those of the principal. The six different headings are used as a way to summarise the factors which will determine whether the balance of costs and benefits is likely to be favourable.

The first heading is the **measure** and its relation to the latent variable (the underlying motivation for the aid contract). A good measure will suffer from little distortion (i.e. there is high alignment between the actions required to increase the latent variable and the performance measure), be a strong indicator of the agent’s actions, be difficult to game and avoid the problems of

Figure 5: Suggested Criterion



asymmetric information (where the agent can forecast the counter-factual more accurately than the principal). It is very rare that PbR should be used in cases with a poor performance measure.

The second heading is the **agent's** characteristics. Many of these affect the costs of PbR relative to other types of aid, for example a risk averse agent with a high discount factor would need a high tariff or would not exert effort in a PbR contract. The degree to which the agent is able to affect the performance measure, and the acceptability of PbR to the agent, is positively correlated with the likely benefit.

The third heading is the **level** of the incentive within the results chain. If the incentive is placed higher up the results chain, there is higher control for the agent but limited potential for efficiency improvements. If the incentive is closer to outcomes or impacts, there is greater room for innovation, but the risk is also higher (meaning a higher tariff must be provided in order to compensate the agent).

The fourth heading is the **tariff**. This is highly context dependent but there are some common elements when deciding upon the correct tariff level. A lower tariff is clearly cheaper for the donor,

and so may be thought of as better value for money. However, there is a risk of non-performance if the tariff is not high enough to incentivise effort. A higher tariff is more able to incentivise effort, but may also lead to greater distortion.

The fifth heading is the ease and cost of **verification**. This is a relatively straightforward cost to predict, but its feasibility should also incorporate the ability to detect gaming and other distortions where possible.

The sixth heading is the **principal's** characteristics. If the donor is unable to commit for a long time horizon, this will affect the type of action that is taken by the recipient, and whether any action is undertaken at all. Furthermore, if the donor is unable to withhold aid in the case of poor performance, any incentive effect is lost.

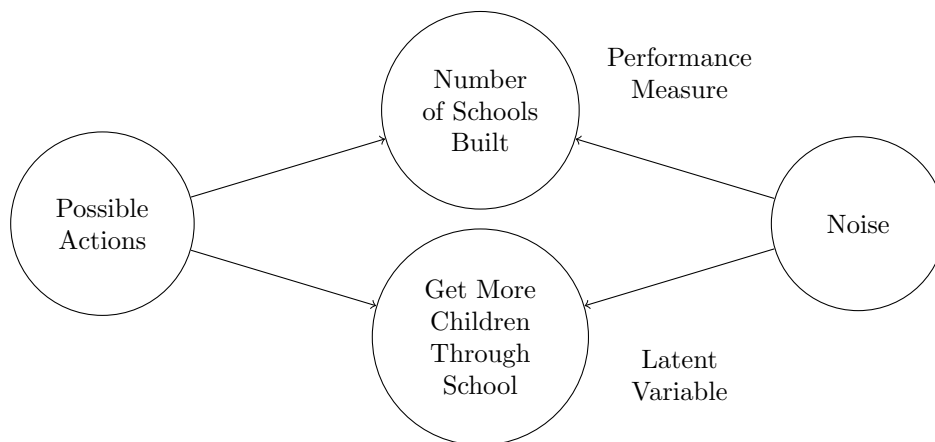
## 6 Worked Examples

Worked examples are provided to give a sense of how the criterion should be applied, and the relevant points used. First, three examples are given based on the level of the measure (output, outcome and impact). In each case only the salient points of the example are discussed: not all points are relevant in every case. All three examples are taken from DFID (2011), and relate to the examples shown in figure 4.

### 6.1 Example: Output

DFID's chosen example for an output is to 'build more schools', suggesting a performance measure of the number of schools or the number of new schools built. Here we provide a worked example examining whether this output measure would be sensible for a PbR contract. We use the number of schools built as the performance measure, though one could think of other precise definitions. Figure 6 displays the problem using the multitask model. For simplicity we assume the true latent variable is the proximate desired effect according to the results chain: increased completion. The relevant factors are discussed below.

Figure 6: Considering Output in a Multitask Model



**Measure** The benefit of transferring the risk is not clear in this case, as the actions of building schools is almost as easily verified as the number of schools built.

Once the best actions to achieve the two goals are considered, the possibility of gaming and distorted incentives become apparent. The measure focuses on the *quantity* of schools, and may thus mean that the size, quality and placement of schools may be affected. In addition, how will the verification be able to distinguish between new school buildings and old buildings? What guarantees the structure will continue to be used as a school in the future?

The signal is much stronger than random noise, as we would not expect a large number of schools to be built in any given year. There is a close relationship between the action and the outcome, and observing the number of schools built will be closely related to the numbers that are attempted to be built.

Asymmetric information is a concern as it is possible that a government was already planning on building schools anyway, and so the results would not be additional.

**Agent** The recipient is likely to have high levels of control. One concern would be whether the agent has the funds to invest in school building in the short run. If this concern is met there are few others regarding the capacity of the agent to build schools.

We would not normally expect a recipient government to be very risk averse in this situation, but the time horizon may be a concern if the government does not expect to remain in power for long enough to reap the benefit.

**Level** It is unlikely that the agent will be able to innovate substantially in the building of schools, but it is possible. The risk level is fairly low, meaning the risk premium should also be low. The time preferences of the agent may have an effect if the time between building and repayment is high.

**Tariff** The tariff will be difficult to calculate, as it is unlikely that there will be a standard cost of building a school - it will depend upon location, size, quality, local soils and so forth. It is easy to imagine a tariff that is too low having no impact, and one that is too high leading to many poor quality schools being built.

**Verification** This should be relatively easy in terms of the measure itself in that buildings are easily verified as present. The difficulty would be in ensuring that the buildings are new and are used as schools.

**Principal** The time horizon is important, as building projects that are well planned will take time. A short time horizon with a generous tariff may lead to schools being built in places where they are not useful. The ability to withhold is important as it may wash away any incentive effects if not present.

## 6.2 Example: Outcome

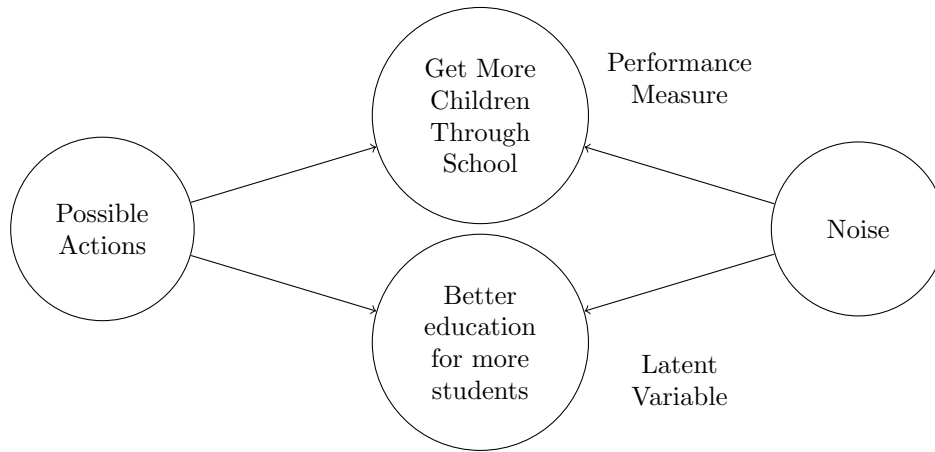
The suggested desired outcome is to ‘get more children through school’. Figure 7 shows the latent variable as a desire to increase the quantity and quality of education provided. In practice the theory of change may justify a focus on completion with a similarly broad goal: in this case the stated reason is the literacy rate.

**Measure** As mentioned elsewhere, the key concern here would be that a measure that simply focuses on completion may lead to quality being neglected.

Asymmetric information is a concern as it is possible that a government was already planning improvements targeted at the completion rate.



Figure 7: Considering Outcome in a Multitask Model



Gaming is possible if students take an end of year exam despite not attending school. Other examples are possible.

The benefit here in transferring risk is obvious - there is enough room for manoeuvre to expect some innovation on behalf of the recipient. The distance between the action and the outcome, and the unobservability of the action are the important considerations.

We would expect completion to have a relatively high signal/noise ratio as the numbers of students enrolled would tend to mean the noise would be relatively small compared to the signal.

**Agent** The control of the agent is likely to be fairly low. Most investments made in increasing completion rates would take a number of years to come to fruition. Because of this, the time preferences of the agent are important: an agent that cannot think long-term would probably not become more efficient.

**Level** The level is such that the agent could be expected to innovate. There is also a fair degree of risk in that some interventions may prove unsuccessful despite the agents best intentions.

**Tariff** The difficulty in setting the tariff will be compounded if a longer time horizon is used as it must steer away from being too high (leading to distortion) and being too low (leading to non performance) for the duration of the project.

**Verification** Verification in education has been widely discussed, see Centre for Global Development (2011).

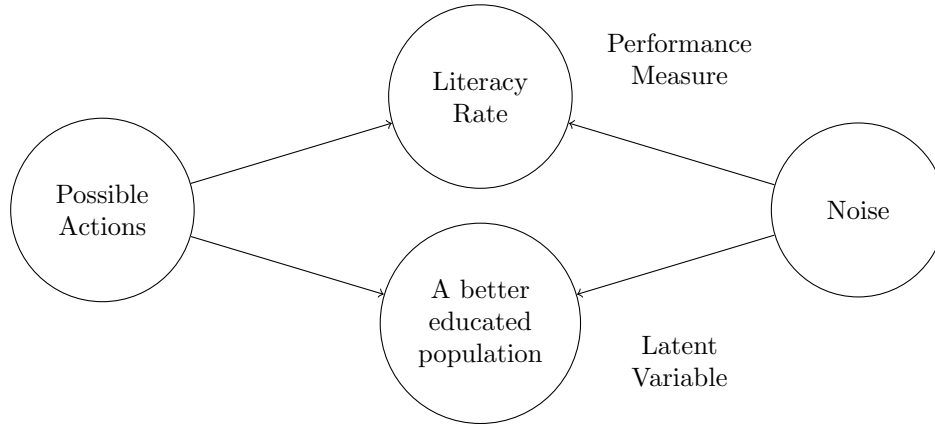
**Principal** The donor is unlikely to be able to commit for 5 years, and so a concern is that more short-term actions may be prioritised over more desirable reforms.

### 6.3 Example: Impact

If the true outcome of interest is a better educated population, the performance measure could be the literacy rate. The latter could be increased in a wide variety of ways. Some would be relatively costly - focusing on children with special needs, say, or on remote areas of the country, or

using holistic approaches to education. Others would be relatively inexpensive: narrowly focusing on what is assessed in literacy tests, in areas with high population density and a pool of underemployed teachers and so forth.

Figure 8: Considering Impact in a Multitask Model



**Measure** A large number of actions that are not easily verified with plenty of discretion for the recipient as to which combination of actions to pursue suggests large potential benefits of transferring the risk to the agent.

A narrow focus on the skills required to do well in literacy tests without addressing functional literacy is conceivable, and so gaming is a concern.

The control of the agent depends upon current levels of literacy. Where literacy rates are very low, actions to improve literacy will be *relatively* easy to detect in the overall measure. Where literacy rates are already high, the agent may be worried that very large campaigns would be needed, and even then may not be recognised within the data.

Distortion is of potentially large concern. An array of actions is required for providing useful and valuable education, whereas only a small subset of these are required for improving literacy: PbR in terms of the literacy rate over-rewards the latter actions, thereby distorting incentives for achieving the unobserved aim of educating the population better.

**Agent Characteristics** Given the likely time between investments and effects, the agent’s characteristics are important.

**Level** The target is at the impact level, meaning that there is room for greater innovation and the accompanying risk is transfer is large. This is also the most innovative and least studied end of PbR.

**Tariff** The tariff would need to be thought generous in comparison with the cost of improving literacy in order to incentivise effort.

**Verification** The cost and feasibility of measuring literacy rates is not prohibitive as this is not an uncommon data gathering exercise.

**Principal** The ability to withhold is crucial here, and this type of target may be better suited to a prize or regional competition rather than one country, given the unpredictability of payment. The ability of the donor to commit for a long enough time period is also questionable.

## 7 Conclusion

Three research gaps have been identified. In each case more theoretical and conceptual work is needed to analyse the situation in a truly satisfactory fashion.

1. In this report we deliberately excluded DIBs as they are not *always guaranteed* to meet the definition of PbR. In many cases DIBs will meet the definition, but further work is required partly due to the flexibility of DIBs. For example, while risk is transferred from the donor to the investor, it is not clear where it will ultimately reside.
2. One rationale for using PbR in cases of high alignment is that PbR can be used as a management tool. Imagine for example that the higher levels of government share priorities with the donor, but lower levels of the recipient (be they civil servants or service providers) do not. It is conceivable in this case that PbR could help the centre drive progress through to the periphery. However, this case is not considered in the conceptual or theoretical literature.
3. Competitions and prizes are closely related to performance based contracts, but not discussed here. This is because they are largely absent from the current debate over performance based contracts. However, they have been discussed in the context of aid (Epstein and Gang, 2009) and are conceptually closely related to performance based contracts. One benefit could be that this could address the difficulty of withholding aid discussed in section 4.1. For example, imagine that aid for education in a variety of countries in a region continues at a fixed rate but that a prize fund is agreed that will be distributed amongst the handful of closely related countries. This may encourage best-practice by creating a discussion around how the results were achieved in successful countries, and create pressure for other countries to emulate the success. A difficulty around this is that aid for a given country currently comes from a given field office’s budget. A number of different field offices would then need to agree to pool resources, possibly reducing the amount of aid that is spent within their country. Alternatively, head office could keep some of its bilateral aid budget unallocated to a specific country.

The criteria and related six features presented in section 5 are the synthesis of the relevant academic and gray literature, and should be read when considering whether PbR is appropriate in a given context. Here, we provide five concluding points related to the conceptual basis for PbR.

1. The potential benefit of PbR lies in the efficiency improvement that comes from linking the agent’s costs with the principal’s utility, in a situation where the agent’s action cannot be observed.
2. The main message of the multitask model is that **the quality of the performance measure** is the determining factor in how strong the incentives should be, and whether a performance based contract should be used at all. In this sense the simple criteria proposed by the Centre for Global Development (“a good measure of progress and a credible way to verify it” 2011, p.vi) are conceptually accurate but deceptively simple. It is not enough for a performance measure to be correlated with the underlying variable of interest *ex ante*, it must also remain so *ex post*.

3. The conceptual evidence implies that donors should be careful about the interpretation of empirical evidence regarding PbR for two reasons. First, neither success nor failure should be generalised too widely. The specific context and related details are likely to determine the effect of PbR to a greater extent than is the case with other types of aid. Second, any apparent success or failure should be substantiated by evidence using something other than the performance measure. If this is not done, an illusory improvement may mask serious detrimental effects.
4. The longer term effects on the types of agent and level of intrinsic motivation are potentially of great importance, and should be monitored closely using data other than the incentivised measure where possible.
5. The principal's characteristics lead to question marks over the viability of the most innovative elements of PbR, where final impacts are incentivised. The room required for true policy innovation is not available if the donor is unable to commit both for long periods and to withhold aid in the case of poor performance.

## References

- Akerlof, G. A. (1970). The market for “lemons”: Quality uncertainty and the market mechanism. *The Quarterly Journal of Economics*, pages 488–500.
- Akerlof, G. A. (1982). Labor contracts as partial gift exchange. *The Quarterly Journal of Economics*, 97(4):543–569.
- Audit Commission (2012). *Local Payment by Results*. available at <http://www.audit-commission.gov.uk/SiteCollectionDocuments/Downloads/20120405localPbR.pdf>.
- Azam, J.-P. and Laffont, J.-J. (2003). Contracting for Aid. *Journal of Development Economics*, 70(1):25–58.
- Baker, G. (2002). Distortion and Risk in Optimal Incentive Contracts. *The Journal of Human Resources*, 37(4):728–751.
- Barder, O. (2009). “Cash On Delivery Aid” - Response to CAFOD. *Mimeo*, <http://www.owen.org/wp-content/uploads/091216-COD-aid-response-to-CAFOD.pdf> last accessed December 2013.
- Barder, O. and Birdsall, N. (2006). Payments for Progress: A Hands-Off Approach to Foreign Aid. *CGD Working Paper*, 102.
- Bernheim, B. D. and Whinston, M. D. (1998). Incomplete Contracts and Strategic Ambiguity. *American Economic Review*, pages 902–932.
- Bertone, M. P. and Meessen, B. (2012). Studying the link between institutions and health system performance: a framework and an illustration with the analysis of two performance-based financing schemes in Burundi. *Health Policy and Planning*, pages 1–11.
- Bevan, G. and Hood, C. (2006). What’s measured is what matters: targets and gaming in the English public health care system. *Public administration*, 84(3):517–538.
- Buchanan, J. M. (1975). *Altruism, morality and economic theory*, chapter The Samaritan’s Dilemma, pages pp. 71–85. New York: Russell Sage Foundation.
- Centre for Global Development (2011). *Cash on Delivery Aid: A New Approach to Foreign Aid Applied to Primary Schooling*. Centre for Global Development: Washington CD.
- Centre for Global Development (2013). Investing in Social Outcomes: Development Impact Bonds. *The Report of the Development Impact Bond Working Group*, available at <http://international.cgdev.org/sites/default/files/investing-in-social-outcomes-development-impact-bonds.pdf> last accessed December 2013.
- Clist, P. (2011). 25 Years of Aid Allocation Practice: Whither Selectivity? *World Development*, 39(10):1724–1734.
- Collier, P. and Dollar, D. (2002). Aidallocation and poverty reduction. *European Economic Review*, 46:1475–1500.
- Collier, P., Guillaumont, P., Guillaumont, S., and Gunning, J. W. (1997). Redesigning conditionality. *World Development*, 25(9):1399–1407.

- Courty, P. and Marschke, G. (2008). A General Test for Distortions in Performance Measures. *The Review of Economics and Statistics*, 90(3):428–441.
- Darley, J. (1991). Setting Standards Seeks Control, Risks Distortion. *Institute of Government Studies Public Affairs Report*, 32(4).
- de la Rosa, L. E. (2011). Overconfidence and Moral Hazard. *Games and Economic Behavior*, 73(2):429–451.
- De Renzio, P. and Woods, N. (2008). The Trouble with Cash on Delivery Aid: A note on its potential effects on recipient country institutions. *Note prepared for CGD initiative on cash on delivery*, pages available at <http://international.cgdev.org/doc/Cash%20on%20Delivery%20AID/Derenzio%20Woods.pdf>, last accessed December 2013.
- DFID (2011). *Guidance on using the revised Logical Framework, How to Note*. DFID: London.
- DFID (2013). DFID Operational Guidance for Payment by Results. *Internal Draft*.
- Dickinson, D. and Villeval, M.-C. (2008). Does monitoring decrease work effort?: The complementarity between agency and crowding-out theories. *Games and Economic Behavior*, 63(1):56–76.
- Easterly, W. (1999). The ghost of financing gap: testing the growth model used in the international financial institutions. *Journal of Development Economics*, 60(2):423–438.
- 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.
- Epstein, G. S. and Gang, I. N. (2009). Good governance and good aid allocation. *Journal of Development Economics*, 89(1):12–18.
- Falk, A. and Kosfeld, M. (2006). The hidden costs of control. *The American Economic Review*, pages 1611–1630.
- Fast, N. and Berg, N. (1975). The Lincoln Electric Company. *Harvard Business School Case*, pages 376–428.
- Gertler, P. and Vermeersch, C. (2012). Using performance incentives to improve health outcomes. *World Bank Policy Research Working Paper*, 6100.
- Gervais, S., Heaton, J. B., and Odean, T. (2011). Overconfidence, Compensation Contracts, and Capital Budgeting. *The Journal of Finance*, 66(5):1735–1777.
- Ghosh, A., Müller, B., Pizer, W., and Wagner, G. (2012). Mobilizing the Private Sector. *CGD Policy Paper*, 15.
- Gibbons, R. (1998). Incentives in Organizations. *The Journal of Economic Perspectives*, 2(4):115–132.
- Gibbons, R. (2005). Incentives between firms (and within). *Management Science*, 51(1):2–17.
- Gibbons, R. (2010). Inside Organizations: Pricing, Politics, and Path-Dependence. *Annual Review of Economics*, 2:337–365.
- Gneezy, U., Meier, S., and Rey-Biel, P. (2011). When and why incentives (don't) work to modify behavior. *The Journal of Economic Perspectives*, 25(4):191–209.

- Gneezy, U. and Rustichini, A. (2000). Pay enough or don't pay at all. *The Quarterly Journal of Economics*, 115(3):791–810.
- Hout, W. (2007). *The politics of aid selectivity: Good governance criteria in World Bank, US and Dutch development assistance*. Routledge: Oxford, UK.
- Koszegi, B. (2013). Behavioral Contract Theory. *Journal of Economic Literature*, Forthcoming.
- Lazear, E. P. (2000). The Power of Incentives. *The American Economic Review*, 90(2):410–414.
- Lazear, E. P. (2003). Teacher incentives. *Swedish Economic Policy Review*, 10:179–214.
- Miller, G. and Singer Babiarz, K. (2013). Pay-for-Performance Incentives in Low- and Middle-Income Country Health Programs. *NBER Working Paper*, 18932.
- Morrissey, O. (2013). Aid Conditionality. *The New Palgrave Dictionary of Economics*, available at [http://www.dictionaryofeconomics.com/article?id=pde2013\\_A000263](http://www.dictionaryofeconomics.com/article?id=pde2013_A000263), last accessed December 2013.
- Mumssen, Y., Johannes, L., and Kumar, G. (2010). *Output-based aid: Lessons learned and best practices*. World Bank: Washington DC.
- Muralidharan, K. and Sundararaman, V. (2011). Teacher Performance Pay: Experimental Evidence from India. *Journal of Political Economy*, 119(1):39–77.
- Neal, D. (2011). The Design of Performance Pay in Education. *NBER Working Paper*, 16710.
- Oxman, A. D. and Fretheim, A. (2008). An overview of research on the effects of results-based financing. *Report Nr 16-2008*. Oslo: Nasjonalt kunnskapssenter for helsetjenesten.
- Savedoff, W. D. (2010). Basic Economics of Results-Based Financing in Health. *Mimeo*, available at <http://www.rbhealth.org/system/files/RBF%20Economics.pdf>, last accessed December 2013.
- Svensson, J. (2000). When is foreign aid policy credible? Aid dependence and conditionality. *Journal of Development Economics*, 61(1):61–84.
- Sylvia, S., Renfu, L., Zhang, L., Shi, Y., Medina, A., and Rozelle, S. (2013). Do you get what you pay for with school-based health programs? Evidence from a child nutrition experiment in rural China. *Economics of Education Review*, 37:1–12.
- Temple, J. R. W. (2010). Aid and Conditionality. *Handbook of Development Economics*, 5:4415–4523.
- Toonen, J., Canavan, A., Vergeer, P., and Elovaino, R. (2009). Learning Lessons on Implementing Performance Based Financing, From a Multi-County Evaluation. *KIT (Dutch Royal Tropical Institute): Amsterdam*.
- Vivid Economics (2013). *Results-Based Financing in the Energy Sector: An Analytical Guide*. Energy Sector Management Assistance Program: World Bank.