
Payment by Results in Development Aid: All That Glitters Is Not Gold

Paul Clist*

Payment by Results (PbR), where aid is disbursed conditional upon progress against a pre-agreed measure, is becoming increasingly important for various donors. There are great hopes that this innovative instrument will focus attention on ultimate outcomes and lead to greater aid effectiveness by passing the delivery risk on to recipients. However, there is very little related empirical evidence, and previous attempts to place it on a sure conceptual footing are rare and incomplete. This article collates and synthesises relevant insights from a wide range of subfields in economics, providing a rich framework with which to analyze Payment by Results. I argue that the domain in which it dominates more traditional forms is relatively small and if it is used too broadly, many of the results it claims are likely to be misleading. The likelihood of illusory gains stems from the difficulty of using a single indicator to simultaneously measure and reward performance: 'once a measure becomes a target, it ceases to be a good measure.' This does not mean PbR should not be used (indeed it will be optimal in some settings), but it does mean that claims of success should be treated with caution. Foreign Aid, Conditionality, Performance, Contract theory, Multitask model. JEL codes: F35, D86, L14

How should development aid be given? Recently, a new answer to this question has been proposed: Payment by Results (PbR), where aid is disbursed contingent upon verified success against pre-agreed measures. While relatively new, PbR resembles several previous attempts to introduce conditionality into aid giving. Similar to traditional conditionality, it defines new and specific goals at the beginning of the agreement. Similar to selectivity, it only disburses aid to recipients that have already met conditions; promises of future actions are not enough. PbR differs because the results already achieved directly determine the amount of aid

The World Bank Research Observer

© The Author 2016. Published by Oxford University Press on behalf of the International Bank for Reconstruction and Development / THE WORLD BANK. All rights reserved. For Permissions, please e-mail: journals.permissions@oup.com
doi:10.1093/wbro/lkw005 Advance Access publication October 5, 2016 31:290–313

disbursed. For example, governments have been paid by donors for each additional child completing primary education or for each vaccination given. Likewise, NGOs have been paid for the number of functioning wells or the additional number of students meeting a set learning outcome.

This new modality has already been adopted by several donors and is being used with recipients of very different sizes, from national governments to small NGOs. The UK's [Department for International Development \(2014a, 12\)](#) states that it is using PbR by default wherever it offers best value for money, with 71 percent of centrally issued contracts having some performance-based element. In 2012, the World Bank approved its Program for Results as its first new modality in 30 years. Other prominent examples include GAVI's use of performance-based funding formulas, the Asian Development Bank's pilot of Results-Based Lending, The Inter-American Development Bank's Results-based Loans, and The European Commission's Variable Results Tranche (see [Perakis and Savedoff 2015](#)).

While there is genuine excitement around the idea, it has rarely been robustly analyzed. Almost every donor that has agreed to some sort of PbR has invented its own terminology, and the plethora of related and overlapping ideas have helped to create a general intellectual fog. However, while the idea is innovative and the empirical evidence base is weak, there is a host of relevant insights from contract theory, behavioral economics, and development economics. Together, these provide a rich framework with which to analyze whether PbR agreements are likely to be more effective than alternatives. These insights relate to three main aspects of a PbR agreement: the agent, the measure, and the principal.

The Principal-Agent Model is the main tool for understanding which characteristics of the agent (aid recipient) determine whether a PbR contract should be used. It focuses on the imbalance between the donor and recipient in terms of priorities (how much the specific goal is shared) and information (as the donor can only observe the recipient from afar). The fundamental question for the donor is whether the PbR contract offered is enough to incentivize the recipient to comply with the donor's wishes, at a price that is still more efficient than a block transfer. How the recipient perceives the contract is thus important and determined by characteristics such as their aversion to risk or loss.

The important features of the measure are best understood using the multitask model, which highlights the fact that any measure upon which performance is contracted will be a proxy. The model illustrates why it is not enough for the measure to be highly correlated with the underlying goal before it is chosen: it must remain so afterwards. This simple insight helps inform donors when PbR contracts will be appropriate. Without this understanding, there is a high likelihood that success will be measured even when it is not present. Even more worryingly, the conditions under which this success will be revealed as "fool's gold" are rare. Good

measures are difficult to find, but they should be a prerequisite for any PbR contract.

The characteristics of the principal (aid donor) have an influence on whether the theoretical models discussed elsewhere resemble the real situation. Development economics has documented donors' experience with conditionality in the past (Svensson 2000; Koeberle 2003) and generally found that donors have found it difficult to impose their own conditions.¹ PbR contracts have legal frameworks that may actually bind donors, but history suggests that these constraints could be sidestepped (e.g., by the donor disbursing the funds using a different aid instrument). The time frame is also a binding constraint for some donors, who will not be able to commit over a long enough period for real results to be achieved. This article proceeds by analyzing the three components of a PbR contract (agent, measure, and principal), discussing it relative to other forms of conditionality, followed by a discussion and a conclusion.

Agent

The Principal-Agent (P-A) model is a useful workhorse for generating and organizing relevant insights from several fields, most notably contract theory (see Azam and Laffont 2003 for a well-known application to the donor-recipient relationship and Savedoff 2010 for an application to results-contingent aid). Imagine a donor (the principal) is about to commit aid to a recipient country or NGO (the agent). The donor can typically choose between a fixed transfer (say General Budget Support or core funding) and a variable PbR contract, where at least some of the aid is disbursed in line with measured performance against some baseline. If the donor and recipient are in complete agreement about how the aid should be spent, a straight transfer is more efficient as it avoids delays in disbursement, uncertain payments, and verification costs. In the P-A model, the fundamental benefit of the PbR contract for the donor is that it changes the incentives for the recipient by linking money to specific measures. This logic underpins conditionality in cash transfers (Das, Do, and Ozler 2005) and salary schemes (Hasnain, Manning, and Pierskalla 2014), where monetary incentives are provided in order to newly incentivize a different set of activities for an individual. In the same way, donors will tend to prefer PbR contracts in cases where they perceive low alignment of preferences.

What Kind of Recipients Suit PbR Contracts?

While the uncertainty of the PbR payment means that it can incentivize different activities, it also means that payments are perceived differently from fixed value contracts. Recipients treat any potential earnings cautiously, discounting them for

risk, volatility, and time. Each of these discounts has an effect on whether the PbR contract is suitable as they lower the effective monetary rewards for success. Say, for example, that a recipient expects a \$10 million transfer with 50 percent chance. If the recipient is risk neutral, it would value that contract at \$5 million. However, most recipients of aid transfers are risk averse, and so the effective value of the contract may be substantially below the \$5 million figure. The more risk averse the agent, the greater the discount. This may have effects on the number and effort of bidders for any contract as the perceived rewards are reduced. In general, smaller agents or organizations will be more risk averse (Miller and Singer Babiarz 2014) as a set transfer will be a larger portion of their budget and their activities will tend to be more concentrated and volatile. In the UK context, PbR contracts have deterred smaller providers from bidding for government contracts (Audit Commission 2012, 16). In this case, the donor could either accept that some recipients will not be incentivized (reducing the number of potential recipients) or increase the promised transfer (reducing value for money).

The recipient's aversion to ambiguity and volatility work in very similar fashion to their risk aversion: a recipient may discount a potential PbR transfer because the chance of it occurring is unknown and may be a volatile source of funds. Barder (2009, 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 defense is within the context of a five- to 10-year agreement with a government based on school completion rates, and in this context the suggested comparison seems fair; ambiguity aversion here should compare volatility in traditional aid to that based on results. While aid volatility is a valid comparison for long-term PbR contracts with governments, it is not clear what the appropriate comparison is for PbR contracts with NGOs or shorter agreements. The nagging 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.

A more nuanced role is played by the recipient's loss aversion (where a recipient weighs losses of a given magnitude more than equal gains) as it simultaneously works in different directions. Loss aversion decreases the relative attractiveness of the contract (compared to a fixed alternative) at the same time as lowering the necessary incentive power. For the former effect, loss aversion would mean that recipients would place a greater weight on situations where effort has been expended but no results have been measured (due to misfortune or problems with the measure). This loss would be felt more keenly, and so a larger reward (in cases of success) would be needed to compensate the recipient. For the latter effect, Koszegi (2014) explains that loss aversion increases the effective power of any given incentive as downswings are felt more intensely. Koszegi uses this insight to explain why two-level incentive contracts (i.e., either a bonus is given or not) are sufficient

to incentivize effort for loss-averse agents. The implication for PbR contracts is that even low-powered incentives (e.g., 5 percent of a budget is subject to performance-related disbursement) could have large effects on behavior as small losses will be magnified.

These theoretical reasons as to why recipients may perceive a PbR contract differently have been seen in various settings. For example, Coffey (2016, 14) reported that several organizations that were initially interested in bidding for work in the education sector as part of the Girls Education Challenge Fund pulled out once they realized the implications of the PbR contract. The reasons given include the need to finance the work up front, with PbR payments subject to the various discounts discussed in this section (risk, loss, volatility, and time). A further reason was simply the time needed to understand the funding mechanism: the rewards on offer did not adequately compensate NGOs for the time cost of understanding the contract. A similar message was found by Ssenooba, McPake, and Palmer (2012), who examined a pilot of performance-based contracting in the Ugandan health sector. They report that the initiative was underfunded: the resources on offer were not enough to incentivize recipients to invest time and resources in understanding and affecting the various performance measures.

Recipients also differ in the extent to which they discount the future, with higher discounts making PbR contracts less attractive. Barder (2009, 6) made the point that PbR contracts with recipient governments take investments that will have positive social returns over the long run and translated them into positive fiscal returns in the short run. The envisaged PbR investments are often attractive investments in and of themselves, but the time preferences of the recipient country are such that the long-term 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 results-contingent 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 costly investments are now needed before benefits are realized. With the standard aid contract, 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 is reversed, and so the aid transfer is time discounted. In other words, agents that are more short-term will tend to find PbR contracts less attractive.

The time dimension is also relevant because agents may be naïve, meaning that they fail to accurately predict their own future preferences or capabilities (de la Rosa 2011; Gervais, Heaton, and Odean 2011). For example, an NGO or government may be overconfident regarding their ability to undertake some action. In this case, the donor would agree to a contract with apparently good terms, only for the recipient to later realize the true situation and not undertake the action. In

cases where a competitive bidding process is allied with an output-based contract, this bias could be prevalent because of the winner's curse. 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 thus reinforces the difficulty of naïveté as the principal incurs a cost as the agent (who overvalued the contract in relation to his costs) may be unable to deliver the contract.

Intrinsic Motivation and the Costs of Control

While the caricatured economist may believe that money is the only motivator and the P-A model the only relevant framework, economics has long recognized that things are not so simple. For example, [Akerlof \(1982\)](#) observed that firms often pay wages that are higher than they absolutely need to and that workers in response often expend more effort than needed. 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. Many of these insights are relevant for PbR contracts in an aid setting. Here, I will discuss strategic ambiguity, intrinsic motivation, and the costs of control.

[Bernheim and Whinston \(1998\)](#) argue that many contracts are less complete than might be expected; a rational decision they term “strategic ambiguity.” The ambiguity in contracts is useful because of two assumptions: (a) performance may be observable but nonverifiable (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) 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 a donor wishes to contract contingent on governance for which there are many possible performance measures, all of which are flawed (as discussed in the next section). All would distort incentives in some way if used as the sole measure (e.g., sham elections could be held to comply in name only, or a complicit auditing firm could be hired). 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, for example, governance. The agent's best response to this situation is to focus on related reforms. 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.

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 that states extrinsic motivation can crowd out intrinsic motivation (see Gneezy, Meier, and Rey-Biel 2011). Gneezy and Rustichini (2000) showed that when small fines were introduced for parents in Israel who were late to pick up their children from day care, the number of parents showing up late increased. This is explained by a shift from intrinsic motivation where arriving late is not fair to 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. The P-A model emphasises that the donor should use PbR contracts more in cases where the recipients undervalue the donor's goals. However, the agent's own goals and preferences may interact with the PbR contract, and so it is likely that a PbR contract may have a lower additional incentive effect for agents who partly share the donor's goals. This emphasises that agents with low alignment with the donor's own goals are more suitable for PbR contracts.

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. However, when 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). With all of the above, there is some way to go before such insights can be applied easily to the PbR setting as most of the empirical work has used individuals rather than organizations. 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.

(Adverse) Selection

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 healthcare 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 health care setting, information asymmetry refers to the fact that individuals may know more about their need for healthcare than the insurance company. This is relevant for PbR in both the short and long run.

In the short run, information asymmetry means that the agent will typically have a more accurate judgement of whether a contract is favorable than the donor. In agreements with national governments, the recipient will then sometimes be rewarded for improvements that would have happened anyway. For example, in a PbR contract with the Rwandan government, a large portion of the money disbursed by DFID for improvements in the numbers of children sitting exams appears to be related to simple demographic cohort effects (Clist 2015). In competitive environments, for example, where a donor is choosing between different NGOs, bidders will tend to have different levels of information regarding the specific context and the profitability of a contract. This could lead to donors rewarding NGOs for having better information on likely trends rather than for delivering improved results.

In the long run, PbR may change the type of agent that bids in competitive environments. In some contexts, this is very positive. In a 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). In short, the best agents (be they individuals, NGOs, or governments) benefit most from moving from a fixed transfer to a system where they get paid for every incremental output as they are able to deliver the highest performance. This process has been observed by Falisse et al. (2014), who looked at a performance-based financing pilot in the health sector in Burundi. Health facilities and workers received income partly on the basis of the number of services offered, with some incentivized services being around 20 percentage points higher than control provinces. Interestingly, provinces that had the performance-based financing model also saw an increase in the number of qualified nurses. While it is not possible to link the change in type of nurse to the improvement, the authors report that qualified nurses heard that they could increase their earnings by moving into PbR provinces. The best nurses saw increases in their salary of around 250 percent over a four-year period. This process, over time, should then see an improvement in the quality of nurses as those able to effectively treat large numbers of patients would be most incentivized to join and remain in the profession.

A caveat to the above argument is that the “best agent” here is defined as one that efficiently produces measured performance, not necessarily one that produces the thing the donor really cares about. In an Indonesian field experiment, Banuri and Keefer (2016) found support for the common idea that greater pay will attract individuals who are less concerned with the donor’s overall aims. It is then possible

that the efficient nurses who are discussed by Falisse et al. (2014) are only more efficient according to measured performance and could neglect nonincentivized or unobservable activities that the donor also cares about. This brings us neatly to the quality of the performance measure, discussed in the Measures section.

Measures

Payment by results puts a great deal of emphasis on the measure and associated tariff as they determine the amount of aid disbursed. Typical tariffs will pay for every unit against some baseline; for example, GAVI disbursed \$20 to eligible governments for each child above a target receiving the DTP3 vaccine in their country (Lim et al. 2008). Because it determines the amount disbursed, the quality of the measure is the single largest determinant of whether a PbR project will be successful. Some measures are demonstrably poor because they do not measure what they are intended to. For example, DFID Ethiopia agreed to tariffs for the number of additional students passing and sitting an exam (Department for International Development 2014b). The inclusion of the “number of passers” was motivated by a desire to incentivize higher standards, but the exam is strictly norm-referenced: a set proportion of sitters should always pass. In essence, the agreed tariff is just paying for all sitters and a proportion of those sitters and cannot incentivize higher standards. While identifying poor measures is quite easy, the definition of a good measure is more technical and relies on insights from a range of fields.

Insights from the P-A Model

The principal-agent model focuses, unsurprisingly, on the characteristics of the principal and agent but does have some insights regarding the measure itself. It distinguishes between actions (which only the recipient can observe) and outcomes (which both parties can observe), recognizing that there is an information asymmetry. For example, a donor may not be able to observe the time and money that an NGO devotes to achieving the donor’s goal of, say, vaccination coverage. They can, however, observe the outcomes by measuring vaccination rates or, ultimately, rates of infection. In the PbR context, there will tend to be a verification cost for observing the outcome as measures take on legal ramifications in many PbR contracts. 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. 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 on a performance-based contract.

The added cost of verification should be balanced against the cost of tracking input costs, assuming that these are reduced in practice.

The credibility of the verification is important as agents will tend to have an incentive to misreport. [Lim et al. \(2008\)](#) and [Sandefur and Glassman \(2015\)](#) present evidence that GAVI's use of a PbR contract (where recipient governments were paid on the basis of the number of vaccinations administered) caused national data systems to over-report their progress. The evidence comes by comparing (non)incentivized data sources for (non)incentivized diseases before and after the GAVI agreement. This kind of triangulation goes far beyond what is currently understood as verification and will typically not be possible in a timely manner as it depends upon the existence of robust and unincentivized data sources. This research, then, raises concerns over the extent to which reported success can be believed in cases where triangulation is not possible: "fool's gold" may be common.

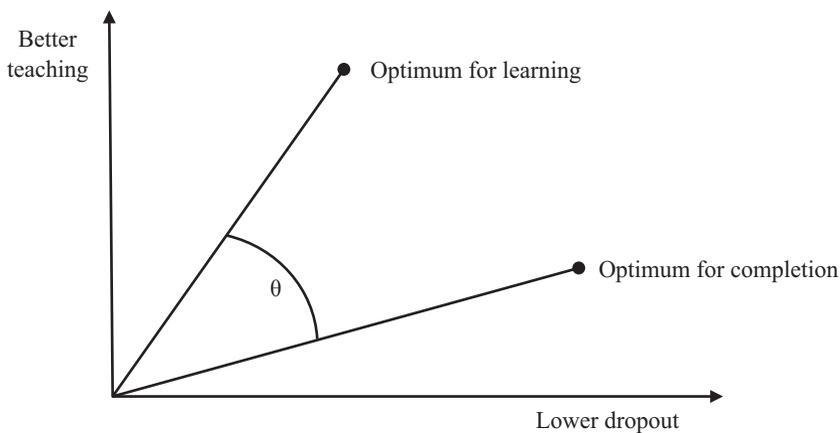
The other major insight taken directly from the P-A model relates to the ratio of signal to noise, or, in other words, the extent to which the recipient can actually affect the outcome. Where the recipient has low control, it will tend to be rewarded due to good luck and punished due to misfortune more often, regardless of their own actions. The ratio of signal (recipient effort) to noise (other factors) then influences the strength of incentive for success and the disincentive for failure. The type of results that are more closely controlled by recipients will tend to be intermediate outputs (e.g., the number of teachers trained) rather than final outcomes (e.g., the number of educated citizens) and tend to be more suitable for PbR (at least in this element).

Multitask Model

The multitask model is a useful tool for thinking about PbR measures from the literature on incentives ([Holmstrom and Milgrom 1991](#)). [Gibbons \(1998, 118\)](#) argues that the performance measure in the P-A model is "misleadingly simple" as it assumes everything that the donor cares about can be summarized in a simple number. The multitask model recognizes that the underlying goals of the donor will seldom be fully captured by any one measure.

Imagine a performance-based contract where a donor agrees to disburse aid to a recipient government based on student completion rates, motivated by the underlying goal of an educated and productive workforce (this example is discussed in a book-length treatment on cash on delivery aid by [Birdsall et al. 2011](#)). As shown in [figure 1](#), the recipient government has two possible actions: he can either increase the quality of education provided or strongly discourage dropout. Intuitively it is obvious that both actions may be desirable, but the balance between the two actions will be different, depending on whether the recipient focuses

Figure 1: The Multitask Model



on the performance measure (completion) or the unobserved true outcome of interest (learning).

Intuitively, one may think that the important consideration with a performance measure is that it is highly correlated with the latent variable, meaning that completion would be a good indicator if it were correlated with learning. However, the multitask model shows us that being correlated before a measure is incentivized is not enough: it must remain so afterwards. In our example, the conceptual problem is that increasing quality may be expensive and have a smaller effect on completion than it does on learning outcomes. Discouraging drop out may be much easier to do, and have a larger effect on completion rates. Therefore, the recipient government could most easily improve completion rates by focusing on dropouts, which may actually harm learning (for example, if disruptive students are kept in classes). More generally, a perfect measure would be perfectly aligned and incentivize exactly the right mix of activities, while the worst measure would only be altered by activities that have no effect on the latent goal. The quality of the measure is then determined by which case it resembles and can be summarized as the alignment between the measures and latent goal, which is given by the angle θ (see [Baker 2002](#) for more).

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*, that is, 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 behavior adapts to incentives and the associated ratio of the two actions. Goodhart's law, cited by [Eldridge and Palmer](#)

(2009, 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 Marchke \(2008\)](#) tested the covariance of performance measures and proxies for the latent variable and found that when a measure is incentivized it has lower covariance with the true latent variable. 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 rewards the right mix of actions even after it is incentivized.

The title of [Kerr's \(1975\)](#) article neatly summarizes 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 PbR element of an aid agreement should be low as weak incentives are more efficient than strong incentives for the wrong action ([Lazear 2000](#)). [Neal \(2011\)](#) takes a more positive line in such situations, arguing that some level of performance-related incentive 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 lack of alignment between the preferences of the principal and the agent.

Gaming and Distorted Incentives

The multitask model provides a framework with which to think about gaming and distinguish it from more complicated unintended consequences. [Baker \(2002, 729\)](#) discusses a good example of gaming from [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 had an influence on the performance measure (the number of keystrokes) but not on the latent variable. We can confidently call this gaming as in this case the action served no useful purpose. However, while gaming receives a lot of attention, there is a different class of unintended effects that are less morally suspect but are probably more important and widespread. These are effects that may not be obvious when considering only a static model or situation and are often unintended and unforeseen consequences.

To illustrate distortion, consider two examples. First, [Sylvia et al. \(2013\)](#) used a randomized controlled trial to study the effects of incentivizing Chinese schools to fight anemia, where control schools were given information and a block grant. Both groups saw anemia rates fall, but incentivized schools also had lower test scores for healthy children. The authors were unsure but thought that the provision of school feeding programs may have negatively affected the amount of teaching time in PbR schools. This situation fits the multitask framework perfectly: the latent goal may have been to improve learning outcomes, but the chosen measure

caused a distortion in the incentives such that school feeding was emphasized. This caused the desired effect in the indicator (anemia), but, as other actions (teaching) were neglected, the end result was actually lower learning outcomes.

Second, [Olken, Onishi, and Wong \(2014\)](#) study a health and education intervention in Indonesian neighborhoods in which groups received either block grants or performance-related disbursements. PbR was significantly more effective than a standard aid transfer in increasing certain indicators over the incentivized 18-month period, but these increases were not present after the program ended (30 months). PbR led to increases in outcomes that were visible in the short run without any improvements that were noticeable in the long run as incentivized activities were front-loaded without any change in the total package of activities. In the multitask framework, this can be understood as a shift toward actions that affect the performance measure in the short run while leaving the latent goal unaffected.

Tensions in the Choice of Measure

There are two main tensions to consider when choosing the measure, and both deal with cases where various desirable features may not be possible simultaneously. The first relates to where “on the results chain” a measure is targeted. A measure could be aimed at ultimate outcomes such as a healthy or educated workforce or at inputs such as numbers of vaccinations administered or children attending school. There is a fundamental tension in this decision that is rarely acknowledged. In order to minimize distortion and maximize recipient discretion, a measure would be targeted at ultimate outcomes. This would make it easier for the measure to capture genuine successes while allowing the recipient the space to innovate. However, an attempt to maximize recipient control by choosing a measure with a low signal-to-noise ratio would lead to a measure of inputs. If inputs are measured, the perceived financial incentive is higher as it limits cases of being rewarded for good fortune and punished for bad luck. These different sets of goals cannot be satisfied simultaneously, and so the design of a PbR contract must trade off concerns over distortion with those over technical capacity and recipient control.

The second tension apparent in the choice of the measure relates to choosing activities for which the recipient has the “right amount” of intrinsic motivation. As argued previously, if the recipient has complete intrinsic motivation (alignment of goals with the donor), the extra PbR contract holds no advantages over a fixed contract. However, if there is not enough intrinsic motivation, results are more likely to be short lived or illusory. To give a micro-example of the temporal aspect, consider evidence from Burundi. [Bertone and Meessen \(2012, 7\)](#) state that over time a bonus that had once incentivized effort became expected regardless of

performance and health workers “reportedly spent time on endless discussions to obtain it, rather than in the provision of services to the population.” The changing perception of the incentive is a cause for concern, if PbR success is to be sustainable. Regarding the possible illusory nature of improvements when recipients have low intrinsic motivation, consider the multitask model introduced earlier. Neal (2011) introduces the idea of intrinsic motivation into this framework, stating that there are norms dictating the amount of time or effort devoted to different actions and that deviation from these is (increasingly) costly. Where intrinsic motivation for the underlying goal of the donor is high, these norms will also be higher, limiting negative distortionary effects. However, if intrinsic motivation is low, then there is greater scope for distortionary effects and illusory measured gains.²

Principal

In the same way that the characteristics of the agent and measure should be taken into account in any analysis of PbR, the principal’s own constraints should be considered. PbR attempts to provide clear rules that govern disbursement while ignoring the pragmatic constraints upon a donor would be naïve, especially as donors do not have a history of enforcing their own rules. Svensson’s (2000) analysis of conditionality argued that the donor’s inequality aversion stymied attempts to withhold aid, and it is not clear that PbR contracts furnish donors with this technology. Furthermore, donors’ own financial systems often make nondisbursal difficult as there are few mechanisms for reallocation of funds. In many cases, the specific PbR contract may not be paid out upon, but if the aid is repurposed for the same recipient using another modality, the spirit of PbR is lost.

Furthermore, the practical difficulty of withholding aid can be seen in the following quote:

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. (Birdsall et al. 2011, 63)

In other words, when factors beyond the recipient’s control affect their performance, 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 percent fall in the world price of a country’s 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 that could conceivably be claimed as mitigating circumstances. In short, the pragmatic considerations that have stymied other attempts to introduce conditionality into aid (e.g., inequality aversion and the difficulty in separating effort from conditions) are not circumvented by PbR.

While it is helpful to analyze a PbR contract in isolation, in the real world a PbR contract cannot be separated from other concurrent or potential aid agreements. While PbR itself may have clear rules on when and whether to disburse, there is much less certainty around other contracts. The measure on which the PbR contract is contingent could also be taken as a barometer of success in that country or with that NGO, and so the rewards for achieving measured success could have wider consequences. This magnifies the perceived incentive to achieve the incentivized goal, with the implication that lower-powered incentives may be sufficient.

Another pragmatic constraint for most donors is the time horizon they are able to contract over, which has potentially large effects on the likelihood of innovation in a PbR contract. [Birdsall et al. \(2011, 20\)](#) suggest a minimum contract period of five years, but the majority of current agreements have a much shorter length ([Holzapfel and Janus, 2015](#)). It is not attractive for donors to remove the flexibility that comes with shorter contracts, and political constraints make this difficult. It is easy to see that when donors are not able to commit themselves for sufficiently long time periods, the likelihood that the recipient undertakes investments that take longer to come to fruition will decrease. This concern becomes binding if we make plausible assumptions regarding the time between an investment and its return. Recall the example from the multitask model where a recipient is being incentivized to increase its completion rates. The two possible actions in the stylized example are either to improve quality or to discourage dropout. The time needed for improvements in teacher quality to lead to increases in completion is high, but the required time for a mechanistic clamp down on dropout and repetition is relatively low. In this case, a donor that agrees to a short-term contract would make it more likely that the worse action would be taken. The constraint for donors over the time horizon seems to point toward more input-based measures being used or to illusory success in outcome measures.

Does PbR Lead to Greater Innovation?

PbR's attractiveness is sometimes fueled by a recognition that donors do not understand how results are best achieved. Proponents of PbR argue that it dominates more traditional forms of aid where the recipient is likely to know (or be able to discover) a more efficient method of achieving a goal ([Perakis and Savedoff 2015](#); [Andrews, Pritchett, and Woolcock 2013](#)). The logic is attractive: donors will typically give project aid with an end goal in mind, and a related set of

activities will be prescribed. Recipients are held accountable for their use of funds against the prescribed activities rather than against the achievement of the goal. As such, where donors prescribe the wrong activities there is a loss in autonomy and, ultimately, efficiency. Such prescriptions would appear to diminish the effectiveness of aid wherever the recipient has greater knowledge than the donor. The counter argument is given by O'Brien and Kanbur (2014, 349), who explain that the World Bank has deliberately chosen to include prescriptive process measures as a way of "transmitting technical expertise." PbR can then be used either on outputs or outcomes, with differing views over the likely innovation and prescriptions needed.

The argument that PbR should be used in order to provide greater innovation is worth examining in greater detail. Essentially there are two assumptions: that PbR will lead to autonomy and that autonomy will lead to innovation. There is some empirical support for the idea that greater autonomy correlates with project success (Honig 2014), but that same research finds that PbR contracts have not led to greater autonomy. Where the counterfactual aid contract is General Budget Support or a block grant, the likely additional autonomy of PbR is not obvious. More generally, it is unclear at this stage whether the lack of an increase in autonomy is due to PbR's current implementation or an inherent feature. Huillery and Seban (2014) offer useful empirical evidence: 96 health areas in the Democratic Republic of Congo were randomly assigned to a receive funding according to either a fixed or results-contingent formula. In keeping with the situation of incomplete alignment discussed above, health workers under the results-contingent treatment expended greater effort. However, this effort was "more of the same" and unsuccessful to the extent that revenue was lower for health facilities under the results-contingent regime. In this particular case, autonomy did not lead to innovation, and the authors speculated that the complicated goals were a factor.

The literature on the optimal incentive contracts for innovation is useful in explaining why PbR contracts do not guarantee innovation, even when autonomy is achieved. Manso (2011) presents evidence consistent with the view that reward for long-term success needs to be combined with short-term tolerance, or even reward, of failure. This early tolerance (or reward) of failure means that agents have the space required to experiment with different strategies, freeing them to discover the most successful approaches. Huillery and Seban (2014) suggest that the health professionals simply did not know how else to provide their services and that the contract did not enable them to experiment in a low-risk setting. Perhaps more surprising are the findings of Ariely et al. (2009), who presented evidence that very high rewards have a detrimental effect on performance for individuals. The psychological effects of large incentives are detrimental, as excessive "arousal" in individuals causes poorer performance. The typical PbR contract does not currently seem to be in danger of having excessively high rewards, but neither does it

resemble the optimum situation of early tolerance for failure. The evidence of [Huillery and Seban \(2014\)](#) seems more typical, with short time horizons that have two negative effects. First, an agent does not have the space to learn and experiment without financial consequences. Second, the reward for discovering the optimal solution is low as the contract is not long enough to recoup any experimentation costs.

An implicit assumption in much of the discussion above is that an unsuccessful project does not represent success for the donor, even if it incurs no costs. Imagine, for example, an overconfident NGO signing a contract that is later found to have tariffs that are too low, and so it simply decides to not try to meet the objectives. In this case, the donor has allocated funds for a given use and so must pay the opportunity costs for the public funds; this is termed the “risk of nonperformance” ([Ghosh et al. 2012](#), 16). It may be that the donor would rather have agreed a more generous contract if it had known the true situation. Furthermore, transfer from principal to agent is never complete as the principal is still liable to suffer reputation risk ([Vivid Economics 2013](#)). It is easy to envisage that the donor’s reputation would suffer because a given result has not been delivered, even if it incurs no direct costs.

Discussion

The preceding analysis has mostly considered PbR relative to block grants, but considering them relative to more traditional forms of conditionality is also informative. There are three main areas in which PbR differs from ex post/ex ante conditionality: the timing of different steps, the ease of enforcing the conditions, and the types of conditions used. First, on the timing of different steps, traditional conditionality starts off contracting an aid recipient by setting out conditions and committing an amount of aid. In theory, these steps are followed by reforms, with aid only disbursed if the donor judges that the reforms were sufficient. In practice, many donors have tended to disburse regardless of whether conditions have been met ([Kanbur 2006](#); [Koeberle 2003](#); [Svensson 2000](#)). The shift towards selectivity can be seen as moving any reform process to before (or outside) the aid contract, with the selection of recipients consisting of a judgement of whether the existing policy and governance environment is suitable. Because of this shift in ordering, there is no reason why commitment and disbursement should differ: ex post conditionality sidesteps the difficulty of enforcing conditions.

It is clear that PbR shares elements with both ex ante and ex post conditionality. At first glance, PbR resembles traditional conditionality because the contracting step comes first. However, whereas traditional conditionality offers a lump sum (e.g., \$5 million, as long as governance reforms are enacted), PbR offers a tariff

(e.g., \$50 per child educated). This means the contracting stage is qualitatively quite different, with PbR offering a clear formula to determine the aid to be disbursed rather than an expected amount. The closest that ex post conditional aid contracts come to PbR is perhaps the Millennium Challenge Corporation's model (Öhler, Nunnenkamp, and Dreher 2012). While there is no formal contract at the outset, the "rules of the game" are clearly set out: to be eligible to receive funds, a recipient must be in the top half of performance in its peer group (judged by 17 independent indicators). This comes close to PbR because recipients are essentially being rewarded for results in a broadly predictable fashion. However, in the MCC model, performance only affects eligibility rather than directly determining the size of aid receipts.

Second, different conditional contracts will differ in how easy they are for the donor to enforce. As noted, ex ante conditional contracts have been difficult for donors to enforce, and ex post conditionality requires no commitment as commitment and disbursement decisions are taken at the same time. The question is, then, whether PbR will be easier to enforce than ex ante conditional contracts, with several reasons why PbR is more easily enforced in at least a narrow sense. PbR is so clear in its tariff, with measured and verified progress, that donors will find it difficult to deviate from the specific disbursement mandated by the measured progress. It is also incremental, which may ease the difficulty of withholding an entire amount of aid in a binary decision on whether reforms have been satisfactorily implemented. However, in a broader sense, it is too early to conclude that PbR allows greater enforcement of conditionality. While PbR contracts may be followed to the letter, the spirit of this conditionality will be undermined if undisbursed funds are reallocated to the same recipient in a different form.

Third, there is variety in the types of conditions that different conditional contracts can use. Ex ante conditionality has traditionally used conditions of many types and forms, from specific conditions over tariff removal to broad assessments of public sector reforms. Smets and Knack (2016) note that the World Bank has shifted to using fewer conditions (and an ex post mindset), but they recognize that there is little research on the type of conditions in current use. Regardless, PbR is by its nature more data intensive as it needs fine-grained data tracking the level of results in a given sector. PbR requires details on things like the specific numbers of school enrolments, children reading to a certain standard, vaccinations, hospital visits, wells built, or police trained. To be useable, this data also needs to be accurate, verifiable and timely. Given current data standards, it is clear that many discussions over whether PbR should be used will be hypothetical as the relevant data simply does not exist.

Where data for PbR conditionality are available, there is an important issue relating to the types of conditions used that is generally neglected. Because PbR focuses on specific and measurable achievements, PbR lends itself to an apolitical

and technocratic approach. Advocates of PbR see this as a strength, arguing that it is well suited to fragile states as it has few preconditions and can reward specific competence (Birdsall et al. 2011). However, PbR will do little to reassure critics who see aid as undermining the rights of the poor by strengthening “benevolent dictators” who are able to deliver specific development results without deep-rooted development (see Temple 2010 and Easterly 2013 for related arguments). While ex post and ex ante conditionality have controversial and contested effects on the general governance context in recipient countries, PbR has an inherent tendency to focus instead on specific achievements and ignore the broader context.

Of the three factors that distinguish PbR from other types of conditional aid, it is this third factor (the types of conditions used and their consequences in terms of policy influence) that is the most important in deciding between modalities. The availability of data will often preclude an attempt to use PbR, and any differences in contract enforceability are currently unproven. However, the types of conditions that can be used are inherent features of different types of aid. The decision of which kind of conditionality to use does not just depend upon the conditions of the recipient country but interacts with the influence and perspective of the donor. The decision will ultimately be based on political economy judgments over likely policy and governance influence, with different donors likely to come to very different decisions.

The Four Ironies of PbR

Having discussed the agent, measure, and principal, this subsection considers PbR in the round by discussing four ironies of PbR. The first irony of PbR is that it is often motivated by a desire to be more data driven, but will tend to diminish the information content of the available data. Success could be measured where there is none (consider the examples of Sandefur and Glassman 2015; Sylvia et al. 2013; Olken, Onishi, and Wong 2014). While the research cited here is able to document this “fool’s gold” effect, in the majority of real-world PbR contracts we will simply never know with certainty whether measured success was genuine.

The second irony of PbR is that citizens’ demands for proof of aid effectiveness may make PbR more popular but ultimately lead to less desirable outcomes. Citizens may justify aid spending in terms of illusory success if measures are susceptible to distortion. Alternatively, they may attribute gains that would have happened anyway to a particular aid intervention. Because PbR contracts only disburse on measured improvements, donors will additionally be able to claim greater value for money. However, wherever there is a mismatch between a given indicator and the latent goals of the donor, such (non-)payments will be misleading.

A third irony of PbR is that, while PbR is often motivated by an acknowledgment that the recipient may have more information than the donor, good PbR contracts may require even more information than a standard block grant. The donor would ideally be aware of the possible distortions of choosing different measures (which requires knowing the mix of actions for both the latent goal and possible measures), the suitability of the time horizon (which requires knowing the likely number of iterations required and the length of the feedback loop), and the correct tariff level. In some situations, PbR will be more demanding for the donor than traditional alternatives.

The fourth irony relates more to aid in general: aid works best where it is needed least (e.g., [Deaton 2013](#)). That is, aid is most effective in countries that already have characteristics such as functioning property rights, a strong legal system, a political ability to manage disputes, and good infrastructure. For aid donors, this creates a quandary as they face a trade-off between aid being “effective” and being “pro-poor.” PbR does not solve this problem but rather reinforces it. As [de Renzio and Woods \(2008, 2–3\)](#) argue, successful countries will receive more aid under PbR, 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?” 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 (i.e., the signal-to-noise ratio is low).

Conclusion

It should now be apparent that Payment by Results does not universally dominate other forms of aid, nor is it universally dominated by them: it is a useful modality that should be used in the right circumstances. The question is, then, the size of the domain in which it should be used and the indicators by which suitability should be judged. This article has discussed many factors, but theoretical insights and empirical experience suggest that the most important prerequisites for PbR success can be summarized in three points: (a) a good and verifiable measure of something continuous that we really care about; (b) a recipient that may undervalue the related improvement, has a reasonable chance of affecting it, and is not excessively worried by the proposed payment structure; and (c) a donor that is able to design and enforce the contract in a reasonable timeframe. These are not universally applicable or guarantors of PbR success, but they do capture the most important indicators of PbR suitability.

While the three characteristics influence the suitability of an individual PbR contract, the measure also determines the quality of the evidence for PbR as a whole. Put simply, bad measures mislead. This article discusses cases where success was measured and led to a disbursement of aid, even when the success appears to have been a temporary illusion (Olken, Onishi, and Wong 2014), spurious (Sandefur and Glassman 2015), or at the expense of the true goal of the donor (Sylvia et al. 2013). If PbR is to be properly judged, evidence of its success will need to be robustly proven from nonincentivized data sources, which are rare. While there will, of course, be cases where PbR has a real advantage over traditional forms of aid, bad measures will lead to a large amount of “fool’s gold” that may be indistinguishable from real success. Positive evidence of PbR should thus be treated with scepticism, and PbR contracts should only be used in the right circumstances.

Notes

*Paul Clist is a Lecturer in Development Economics in the School of International Development, University of East Anglia; email: paul.clist@uea.ac.uk. This article is loosely based on “The conceptual basis for Payment by Results in International Development” (Clist and Verschoor 2014), which I’m happy to acknowledge was funded by the Department for International Development (DFID UK). The views expressed do not necessarily reflect the UK Government’s official policies. I have benefited from comments from and conversations with Ellie Cockburn, Stefan Dercon, Roger Drew, Donald Menzies, Rita Perakis, Fadia Saadah, William Savedoff, Pieter Serneels, Arjan Verschoor, and Brendan Whitty. I would also like to thank attendants at presentations at the World Bank, Washington DC, and DFID UK. All views, and mistakes, remain my own.

1. While more traditional forms of conditionality have often discussed fungibility, Payment by Results schemes do not typically ring-fence the transfer and so fungibility is not strictly relevant. For example, Birdsall et al. (2011) do not mention fungibility at all in their book-length treatment of PbR, instead seeing the payment as a reward or compensation for results achieved.

2. In the health domain, Gertler and Vermeersch (2012) argue that there are ethical limits of care that limit distortionary effects. In the education domain, 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. Both sets of authors thus argue that there is limited scope for distortion and view PbR contracts in a positive light.

References

- Akerlof, G. A. 1982. “Labor Contracts as Partial Gift Exchange.” *The Quarterly Journal of Economics* 97 (4): 543–69.
- Andrews, M., L. Pritchett, and M. Woolcock. 2013. “Escaping Capability Traps through Problem Driven Iterative Adaptation (PDIA).” *World Development* 51: 234–44.

- Ariely, D., U. Gneezy, G. Loewenstein, and N. Mazar. 2009. "Large Stakes and Big Mistakes," *Review of Economic Studies* 76: 451–69.
- Audit Commission. 2012. *Local Payment By Results*. London: Audit Commission.
- Azam, J. P., and J. J. Laffont. 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): 728–51.
- Banuri, S., and P. Keefer. 2016. "Pro-Social Motivation, Effort and the Call to Public Service." *European Economic Review* 83: 139–64.
- Barder, O. 2009. "Cash on Delivery Aid - Response to CAFOD." Mimeo. Accessed March 2016. <http://www.owen.org/wp-content/uploads/091216-cod-aid-response-to-cafod.pdf>.
- Berg, N. A. 1975. "The Lincoln Electric Company." *Harvard Business School Case 376-028*. Harvard University, Boston, MA.
- Bernheim, B. D., and M. D. Whinston. 1998. "Incomplete Contracts and Strategic Ambiguity." *American Economic Review* 88 (4): 902–32.
- Bertone, M. P., and B. Meessen. 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* 28 (8): 1–11.
- Birdsall, N., A. Mahgoub, W. Savedoff, and K. Vyborny. 2011. *Cash on Delivery Aid: A New Approach to Foreign Aid Applied to Primary Schooling*. Washington, DC: Centre For Global Development.
- Clist, P., and A. Verschoor. 2014. "The Conceptual Basis of Payment By Results." Mimeo. Accessed March 2016. <http://r4d.dfid.gov.uk/output/195131/>.
- Clist, P. 2015. "Evaluation of Results Based Aid in Rwandan Education." Econometric Report. Norwich: Mimeo.
- Coffey. 2016. "GEC Process Review Report." Mimeo. Accessed March 2016. https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/501596/process-review-report.pdf.
- Courty, P., and G. Marschke. 2008. "A General Test for Distortions in Performance Measures." *The Review of Economics and Statistics* 90 (3): 428–41.
- Das, J., Q. T. Do, and B. Ozler. 2005. "Reassessing Conditional Cash Transfer Programs." *The World Bank Research Observer* 20 (1): 57–80.
- de la Rosa, L. E. 2011. "Overconfidence and Moral Hazard." *Games and Economic Behavior* 73 (2): 429–51.
- de Renzio, P., and N. Woods. 2008. "The Trouble with Cash on Delivery Aid: A Note on its Potential Effects on Recipient Country Institutions." Mimeo. Accessed March 2016. <http://international.cgdev.org/doc/cash%20on%20delivery%20aid/derenzio%20woods.pdf>.
- Deaton, A. 2013. *The Great Escape: Health, Wealth, and the Origins of Inequality*. Princeton, NJ: Princeton University Press.
- Department for International Development. 2014a. "DFID's Strategy For Payment By Results: Sharpening Incentives to Perform." Policy Paper. London: DFID.
- Department for International Development. 2014b. "Business Case and Summary 202989." Accessed March 2016. <https://devtracker.dfid.gov.uk/projects/gb-1-202989/documents>.
- Dickinson, D., and M. C. Villeval. 2008. "Does Monitoring Decrease Work Effort?: The Complementarity between Agency and Crowding-Out Theories." *Games and Economic Behavior* 63 (1): 56–76.

- Easterly W. 2013. *The Tyranny of Experts: Economists, Dictators, and the Forgotten Rights of the Poor*. New York: Basic Books.
- Eldridge, C., and N. Palmer. 2009. "Performance-Based Payment: Some Reflections on the Discourse, Evidence and Unanswered Questions." *Health Policy and Planning* 24 (3): 160–66.
- Falisse, J. B., J. Ndayishimiye, V. Kamenyero, and M. Bossuyt. 2014. "Performance-Based Financing in the Context of Selective Free Health-Care: An Evaluation of its Effects on the Use of Primary Health-Care Services in Burundi Using Routine Data." *Health Policy and Planning* 30 (10): 1,251–60.
- Falk, A., and M. Kosfeld. 2006. "The Hidden Costs of Control." *The American Economic Review* 96 (5): 1,611–30.
- Gertler, P., and C. Vermeersch. 2012. "Using Performance Incentives to Improve Health Outcomes." *Policy Research Working Paper 6100*. World Bank, Policy Research Department, Washington, DC.
- Gervais, S., J. B. Heaton, and T. Odean. 2011. "Overconfidence, Compensation Contracts, and Capital Budgeting." *The Journal of Finance* 66 (5): 1,735–77.
- Ghosh, A., B. Müller, W. Pizer, and G. Wagner. 2012. "Mobilizing the Private Sector." Policy Paper 15. Washington DC: Center for Global Development.
- Gibbons, R. 1998. "Incentives in Organizations." *The Journal of Economic Perspectives* 2 (4): 115–32.
- Gneezy, U., S. Meier, and P. Rey-Biel. 2011. "When and Why Incentives (Don't) Work to Modify Behaviour." *The Journal of Economic Perspectives* 25 (4): 191–209.
- Gneezy, U., and A. Rustichini. 2000. "Pay Enough Or Don't Pay at All." *The Quarterly Journal of Economics* 115 (3): 791–810.
- Hasnain, Z., N. Manning, and J. H. Pierskalla. 2014. "The Promise of Performance Pay? Reasons for Caution in Policy Prescriptions in the Core Civil Service." *The World Bank Research Observer* 29 (2): 235–64.
- Holmstrom, B., and P. Milgrom. 1991. "Multitask Principal-Agent Analyses: Incentive Contracts, Asset Ownership, and Job Design." *Journal of Law, Economics, and Organization* 7: 24–52.
- Holzapfel, S., and H. Janus. 2015. "Improving Education Outcomes by Linking Payments to Results." Discussion Paper 2015/2. Bonn: DIE.
- Honig, D. 2014. "Navigation by Judgment: Organizational Autonomy in the Delivery of Foreign Aid," Harvard University, Mimeo, Boston, MA.
- Huillery, E., and J. Seban. 2014. "Performance-Based Financing, Motivation and Final Output in the Health Sector: Experimental Evidence from the Democratic Republic of Congo." Discussion Paper 2014-12. Paris: Sciences Po Economics.
- Kanbur, R. 2006. "The Economics of International Aid." In Serge-Christophe Kolm and Jean Mercier Ythier, eds., *The Handbook of the Economics of Giving, Altruism and Reciprocity*, Volume 2. Elsevier.
- Kerr, S. 1975. "On the Folly of Rewarding A, While Hoping For B." *The Academy of Management Journal* 18 (4): 769–83.
- Koerberle, S. G. 2003. "Should Policy-Based Lending Still Involve Conditionality?" *The World Bank Research Observer* 18 (2): 249–73.
- Koszegi, B. 2014. "Behavioral Contract Theory," *Journal of Economic Literature* 52 (4): 1,075–118.
- Lazear, E. P. 2000. "The Power of Incentives," *The American Economic Review* 90 (2): 410–14.
- 2003. "Teacher Incentives." *Swedish Economic Policy Review* 10: 179–214.

- Lim, S., D. Stein, A. Charrow, and C. Murray. 2008. "Tracking Progress towards Universal Childhood Immunisation and the Impact of Global Initiatives: A Systematic Analysis of Three-Dose Diphtheria, Tetanus, and Pertussis Immunisation Coverage," *The Lancet* 372: 2,031–46.
- Manso, G. 2011. "Motivating Innovation." *The Journal of Finance* LXVI (5): 1,823–60.
- Miller, G., and K.S. Babiarz. 2014. "Pay-for-Performance Incentives in Low- and Middle-Income Country Health Programs," *Encyclopedia of Health Economics*, 457–466.
- . 2014. "Pay-for-Performance Incentives in Low- and Middle-Income Country Health Programs," *Encyclopedia of Health Economics*, 457–466.
- Muralidharan, K., and V. Sundararaman. 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," *Handbook of Economics of Education*. Volume 4, 495-550. <http://dx.doi.org/10.1016/B978-0-444-53444-6.00006-7>
- O'Brien, T., and R. Kanbur, R. 2014. "The Operational Dimensions of Results-Based Financing." *Public Administration and Development* 34: 345–58.
- Öhler, H., P. Nunnenkamp, and A. Dreher. 2012. "Does Conditionality Work? A Test for an Innovative US Aid Scheme." *European Economic Review* 56 (1): 138–53.
- Olken, B.A., J. Onishi, and S. Wong. 2014. "Should Aid Reward Performance? Evidence from a Field Experiment on Health and Education in Indonesia." *American Economic Journal: Applied Economics* 6 (4): 1–34.
- Perakis, R., and W. Savedof. 2015. "Does Results-Based Aid Change Anything? Pecuniary Interests, Attention, Accountability and Discretion in Four Case Studies." Policy Paper 52. Washington DC: Center for Global Development.
- Savedoff, W. D. 2010. "Basic Economics of Results-Based Financing in Health." Mimeo. Accessed March 2016. <http://www.focusintl.com/rbm082-rbf%20economics.pdf>.
- Sandefur, J., and A. Glassman. 2015. "The Political Economy of Bad Data: Evidence from African Survey and Administrative Studies." *Journal of Development Studies* 51 (2): 116–32.
- Smets, L., and S. Knack. 2016. "World Bank Lending and the Quality of Economic Policy." *The Journal of Development Studies* 52 (1): 72–91.
- Ssenooga, F., B. McPake, and N. Palmer. 2012. "Why Performance-Based Contracting Failed in Uganda—An 'Open-Box' Evaluation of a Complex Health System Intervention." *Social Science and Medicine* 75 (2): 377–83.
- Svensson, J. 2000. "When Is Foreign Aid Policy Credible? Aid Dependence and Conditionality." *Journal of Development Economics* 61 (1): 61–84.
- Sylvia, S., L. Renfu, L. Zhang, Y. Shi, A. Medina, and S. Rozelle. 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." In Dani Rodrik and Rosenzweig Mark, eds., *The Handbook of Development Economics*, 5: 4415–523.
- Vivid Economics. 2013. *Results-Based Financing in the Energy Sector: An Analytical Guide*. Washington, DC: Energy Sector Management Assistance Program, World Bank.