Centre for Understanding Behaviour Change

www.cubec.org.uk

Payment by Results in Early Years Provision

Haroon Chowdry

July 2011

Short Policy Report No. 11/03

(Funded by Department for Education)

Centre for Understanding Behaviour Change
Centre for Market and Public Organisation
Bristol Institute of Public Affairs
University of Bristol
2 Priory Road
Bristol
BS8 1TX
UK

www.cubec.org.uk

CUBeC delivers evidence and insight into the drivers of behaviour change to inform and improve policy-making. The Centre combines expertise in a wide range of academic disciplines: economics, psychology, neuroscience, sociology, education, and social research.

Payment by Results in Early Years Provision

Literature review

1. Introduction

This document is a literature review commissioned by the Department for Education into the role for, and implications of, payment by results in public service delivery. The Department is exploring the potential usage of such a system in the context of the early years services, funded from within the Early Intervention Grant. However, the Department has noted an evidence gap on the usage of payment by results systems. This document aims to help inform any future policy decisions by exploring in the rationale for implementing payment by results, and the relevant empirical evidence from applications of payment by results across a range of countries and services.

This document is set out as follows: Section 2 sets out the high-level theoretical and political rationale for payment by results, then Section 3 focuses in detail on specific theoretical insights from behavioural economics. Section 4 then summarises the lessons to be learned from previous applications of financial incentives in the delivery of public services. Section 5 concludes.

2. What is the rationale for payment by results?

2.1. Theory

The theoretical justification for payment by results (PBR) stems from a well-known model in economics (Ross, 1973) whereby an employer ('principal') enters into a contractual agreement with a worker ('agent'). The principal's objectives are to optimise some measure of final outcomes, such as maximising the level of output or minimising the costs of providing a service. The agent may not share these objectives, and may care solely about his or her wage. Furthermore, there is a psychological cost to the agent of putting in effort. The prediction from the theory is that if the principal cannot perfectly observe the agent's effort ('imperfect information') *and* if the two parties' objectives are not aligned then the agent will shirk, resulting in worse outcomes for the principal. To help rectify this, the principal may specify the contract to pay more to the agent if output (which can be observed) is higher, and less if it is lower. This then re-aligns the agent's with the principal's and should provide the agent with a net incentive to put in more of an effort. If investing more effort does indeed improve output, then the principal achieves a better final outcome. Making payments on the basis of outcomes is therefore justified on efficiency grounds.

Dixit (2002) enumerates three types of informational asymmetry that could be invoked as describing the relationship between the principal and the agent, and as justification for a PBR system. The first is moral hazard, where the action chosen by the agent (effort) takes place after the agreement is made. The second is adverse selection, where the agent has some private information about their own productivity or ability before the contract with the principal is signed; the task for the principal is to then design a scheme that induces the agent to reveal this information. The final asymmetry is known as costly verification, where the issue is instead that the

agent can observe the outcome better than the principal can. In this case, the principal must devise a remuneration scheme involving some method of verifying the outcome, such as an audit, which may involve additional costs. Of these three scenarios, the moral hazard interpretation is the most common one when considering incentives in public sector organisations, as it has the most relevance to the issues that arise in a practical context.

Nevertheless, as Burgess & Ratto (2003) point out, standard principal-agent theory has numerous simplifying features that might reduce its applicability to the public sector. First is the assumption of one principal; in the public sector, employees may be accountable to several layers of management and organisations, and are ultimately accountable to taxpayers as well. Next is the assumption of one agent; individuals often collaborate in departments or teams, which introduces its own nuances, while services can be commissioned to a range of possible providers who must compete with one another. Thirdly, the standard theory assumes that the agent knows how to improve outcomes and can simply do so by exerting more effort. This idea may be less applicable in public services, where there is not only uncertainty over the precise methods that improve outcomes, and where services are also often delivered in challenging circumstances to disadvantaged people, whose outcomes might be more difficult to influence. There is an assumption of measurability: the principal can perfectly observe the (singular) outcome that matters. In a public services environment, however, there may be multiple outcomes, some more measurable than others. Quality, for example, is more difficult to observe than the quantity of public services, but is also important. If quantity of output is targeted in an incentive scheme purely that is what can be measured, then quality may well be neglected as agents focus on maximising the outcomes that the incentive scheme targets.

There is also the issue of intrinsic motivation: the innate pleasure or pride that employees or agents may take from doing a job. This may be particularly relevant for the public sector. If agents are intrinsically motivated to do their work then they may not be particularly responsive to financial (extrinsic) incentives. Worse still, intrinsic motivation might actually suffer when financial incentives are introduced, leading to a theoretically ambiguous effect on overall performance. This is discussed further below.

These numerous and important caveats raise questions about the applicability of incentive scheme in the public sector. Heckman et al. (1997) neatly sum up the relevance of performance measures as depending on whether (i) agents have specified goals, (ii) agents' goals can be quantified in order to determine success or failure. These conditions tend to be satisfied in settings where there are close private sector substitutes for the public service activity in question. This may or may not be true in the case of early years provision.

2.2. Policy

Theoretical considerations aside, performance management in the public sector may still be attractive to policymakers for more practical reasons. Clearly the offer of financial incentives contingent on performance transfers risk from policymakers/commissioners to agents/providers, which may be desirable from the perspective of the former. However, regardless of whether formal incentive schemes are implemented, the act of defining measures of performance itself can lead to service improvements – even if the actual process required to improve performance is unknown

(Propper & Wilson, 2003). They add that performance management can foster accountability and generate pseudo-competition between agents, both of which may be regarded as beneficial.

Smith (1990) identifies further benefits of measuring performance in the public sector. It can help clarify an organisation's objectives and serve as an information system for the managers of both that organisation and for the managers in a commissioning authority. Also, such indicators can be used to create a standards framework which may be useful for policy evaluation and formation.

3. Issues arising from behavioural economics

This section sets out some of the key areas of psychology and behavioural economics that PBR invokes.

3.1. Framing

Framing is a cognitive bias, a failure of standard rational choice. It occurs when the same informational about uncertain outcomes in different states of the world, presented ("framed") in different ways, induces people to make different choices or reverse their preferences. The seminal reference for this is Tversky & Kahneman's (1981, 1986) experiment involving a fictitious 'Asian disease problem'. Participants were told to imagine that their country faced an outbreak of disease from Asia, which could result in up to 600 deaths. They were told to choose between two treatments: A, which resulted in a certain outcomes, and B, which resulted in uncertain outcomes.

When treatment A was framed as saving 200 lives with certainty, and B framed as saving 600 people with probability 1/3 and nobody with probability 2/3, participants were more likely to prefer A – indicating that they were risk-averse. However, when A was framed as resulting in 400 deaths with certainty, and B framed as resulting in zero deaths with probability 1/3 and 600 deaths with probability 2/3, participants were more likely to prefer B – indicating that they were risk-loving. This is was purely a consequence of presenting the same information in a different way. Hence people tend be risk-averse when uncertainty is described in terms of potential gains, and risk-loving when uncertainty is described in terms of potential losses.

If people react differently to uncertainty depending on whether it is described as potential gain or a loss, this may have implications for incentive schemes where risk is introduced and transferred to agents/providers. This has been considered in the literature. Conrad & Perry (2009) argue that incentive schemes framed in terms of rewards or bonuses are less effective than those framed in terms of penalties, because of risk-aversion. Agents may not care much about potential gains but will be very keen to avoid potential losses. Van de Weghe & Bruggeman (2006) concur with this finding (and also claim that using multiple measures of performance increases the effort expended).

On the other hand, programmes framed in terms of bonuses are more likely to be received and perceived favourably, which may be an important consideration. Luft (1994) argues that bonus schemes connote notions of approval and reward (compared to schemes featuring penalties). Also, schemes involving bonuses for better performance involve an implicit basic/guaranteed remuneration, which might be valued by agents/providers. Thus a negatively-framed contract

(involving penalties rather than bonuses) might have unintended consequences for morale or motivation.

Overall, the consequences of framing biases are not unambiguous: arguments exist for both positively- or negatively-framed contracts leading to improvements in performance. Where there is disagreement in the literature, it may be due to different kinds of framing being conflated with one another. Levin et al. (1998) identify three separate kinds of framing: the first is risky-choice framing, such as the Asian disease problem above. The second is attribute framing, where people assess a product or situation differently if it is described differently despite being unchanged – importantly, this has nothing to do with risk or risk aversion. It is caused by people's sensitivity to the specific language used to describe a situation where no uncertainty is involved (such as the attributes of a product in a supermarket). Finally, there is goal framing, where people's tendency to do something depends on whether they are told the benefits of doing it, or the risks of not doing it. This is also subtly different: there is no a preference reversal as both messages will induce people to do the activity; instead, one type of message – typically the negatively-framed one – may be more effective than the other. An example of this is breast self-examination among women who may be at risk of breast cancer. Emphasising the benefits of breast self-examination (in terms of early detection and higher recovery rates) has been found less effective than emphasising the costs of not doing self-examination (late detection and lower recovery rates). These sorts of framing can all arise independently of one another, and should all be borne in mind when considering possible psychological responses to uncertainty, language and terminology.

3.2. Motivation and effort

Understanding the impact of incentives relies on understanding an agent's motives to begin with. Incentives that are related to factors which motivate performance are clearly likely to be more effective at increasing performance. As mentioned in Section 2.1, workers and service providers in the public sector may have intrinsic motivation, deriving some satisfaction from the delivery of public services. If agents are intrinsically motivated to do their work then they may not be particularly responsive to financial (extrinsic) incentives.

Worse still, intrinsic motivation might actually suffer when financial incentives are introduced, leading to a theoretically ambiguous effect on overall performance. Recent work in experimental economics has revealed three examples in the literature. Gneezy & Rustichini (2000) found that, while stronger financial incentives generally increase performance, financial incentives are altogether less effective than systems of fixed remuneration, because they lead to reductions in intrinsic motivation.

Irlenbusch & Sliwka (2005) studied an experiment where remuneration for a task could be a mixture of fixed payments and piece rates. They found that if principals had the freedom to offer piece rates, agents reduced their effort levels making both parties worse off. However, reinstating fixed wages reduced efforts even further. The authors concluded that while extrinsic incentives can crowd out intrinsic ones, agents' past experiences of compensation schemes are important interacting factor. Fixed payments are not a superior method of compensation unless they were previously in place.

Wiebel et al. (2009) also find that payment on the basis of performance causes a cognitive shift in which higher extrinsic motivation is offset by lower intrinsic motivation. While the overall effect on performance is unclear, the authors note that outcome-contingent payments are more likely to be effective for activities that are monotonous or less interesting. Not only do such activities carry less intrinsic value to begin with, but output from such tasks is easier to measure and is easier to attribute to a particular individual.

On the other hand, research by Sprinkle (2000) found that financial incentives can improve performance. In an experiment with fixed or incentive-based remuneration, it was found that in the latter case, participants devoted more time to completing tasks and made more use of the information that was available. However, these performance gains only materialised in later rounds of the experiment, after participants had acquired experience and feedback on their performance. This suggests that incentive schemes may require a certain amount of learning on the part of agents in order to be effective. Moreover, the experiment in this study used undergraduate business students, so its findings may not be generalisable (especially to a public sector one).

3.3. Size of financial incentives

The psychological literature has frequently covered areas around payment for performance between employers and employees; see Rynes et al. (2005) for a recent review. The basic message from their work is that money clearly matters and has considerable motivational implications, but only some of these are desirable. Thus the challenge for commissioners of services is to ensure that the exact specification of outcome-contingent contracts minimises unintended or negative motivational behaviour.

It is widely accepted that stronger financial incentives are likely to increase performance. However they may not be preferable from the view of employees or service providers. Individuals may be risk-averse (a common assumption in economics and psychology) in which case they are likely to prefer contracts with smaller incentives: these provide more stable remuneration and allow them to 'insure' against eventualities where productivity is low or where targets are not reached. Another issue with strong financial incentives is that they may exacerbate existing concerns about imperfect performance measures. It is accepted that payment-by-results can incentivise agents to redirect their effort towards only outcomes that are measurable, or otherwise manipulate the reported outcomes; stronger incentives magnify this risk. They should therefore be seen as a potentially high-risk, high-return strategy (Gerhart et al., 1996).

Gneezy & Rustichini (2000) studied the effect of varying financial incentives in an experimental setting. They found that, given the existence of a positive financial incentive, offering more money resulted in better performance than offering less money. However, participants performed better still in activities which had no financial incentive at all. The researchers try to rationalise this with the idea that intrinsic motivation suffers when extrinsic incentives are introduced. They conjecture that activities normally considered noble (or a duty) could be reduced when monetary compensation is introduced, as it can undermine the sense of civic duty or pride in the activity. This clearly has implications for payment on the basis of outcomes in public sector environments. The issue here is not the remuneration itself, but the fact that it is outcome-contingent: it is claimed

that a fixed level of remuneration (independent of performance) does not lead to reductions in intrinsic motivation.

The findings from the literature do not pinpoint an optimal intensity of financial incentives or optimal proportion of remuneration that should be outcome contingent. However, there are guiding principles that might assist in making decisions about this. Milgrom & Roberts (1992) suggest that the optimal intensity of incentives depends on four factors: (i) the responsiveness of employees' effort levels to the incentive; (ii) the extent to better outcomes are driven by more effort (rather than luck or uncertainty); (iii) the ease with which outcomes of value can be measured; (iv) the amount of risk employees and service providers will accept. There are clearly trade-offs involved in weighing these (sometimes conflicting) criteria against each other.

4. Insights and evidence from policy experience

Outcome contingent payments have a long history of being used in the delivery of public services, both in the UK and abroad. This section considers the some general concerns that have been raised in research on payment for performance in the public sector, then draws out the main findings and lessons that can be learnt from the application of performance incentive systems in other areas of public policy.

4.1. General issues

Concerns have been raised in the research about unintended consequences that have arisen from certain applications of payment by results: in particular, their failure to achieve a desired improvement in outcomes, and their possible negative effects on morale among workers.

It has been observed in the literature that performance indicators and actual performance are not highly correlated with each other (Van Theil & Leeuw, 2002). Heckman et al. (2002), in their analysis of the Job Training Partnership Act in the US, concur with this idea. Why might measuring and rewarding observed performance fail to create a significant increase in performance itself? Eldridge & Palmer (2009) cite Goodhart's Law – "When a measure becomes a target, it ceases to be a good measure" – to draw attention to the risk that incentive schemes, manipulated to produce an increase in the target output without adding increased value to the services rendered. Smith (1993) lists seven ways in this inefficiency may arise:

- Tunnel vision – concentrating on activities targeted by a performance measure, to the detriment of other valuable activities whose outcomes are not measured;
- Sub-optimisation – taking a narrow view of the aims an organisation, rather than a higher-level strategic view of its long-term objectives;
- Myopia – since outcomes are likely to be measured in the short-term (e.g. on an annual basis), they incentive activities that yield immediate benefits rather than investments which may achieve (possibly larger) returns in the long term;

- Convergence – targets and performance indicators may be defined on the basis of other service providers' performance. This can, theoretically, create incentives for 'standard' or 'average' outcomes rather than performance which would result in being an outlier.
- Ossification – target-driven activities may stifle innovative or experimental activities as they will be seen as risky.
- Gaming – manipulation of some form that allows targets to be reached in a way that does not challenge the organisation to actually perform better (such as moderate performance when baseline measures are collected).
- Misrepresentation – distorting reported performance outcomes without any change in actual behaviour, perhaps via falsification of data on performance.

So far this has focussed on risks that organisation acting rationally may face. However, there may also be consequences of a more human nature. The way individual workers perceive or respond to an incentive scheme is clearly important as it may determine the eventual output response. Surveys of civil service workers facing performance-related-pay schemes have suggested that such schemes carry a considerable risk of negative effects on workers' motivation and perceptions.

Marsden & Richardson (1992) conduct a detailed survey of staff at the then Inland Revenue, to gauge their views on final incentives. These views were generally mixed: while a majority of staff supported the *principle* of performance-related pay, a significant number were hostile to it, and many perceived it to be unfair. Moreover, many participants felt that such a scheme would not be effective, firstly because the incentives in the scheme were not large enough in magnitude to elicit a higher productivity level, and secondly because they felt they were already working to the best of their ability. The authors conclude that the scheme may have resulted in a net reduction in workers' motivation levels.

Such messages are corroborated in more recent work by Marsden & French (1998), who looked at a range of public service organisations including schools and hospitals. They found that while staff generally supported the idea of rewarding higher performance, most interviewees did not believe performance-related pay raised their motivation. Worse still, many people believed it was a divisive policy that made other colleagues jealous – especially if it was deemed to have been combined with managerial favouritism.

4.2. Insights from welfare-to-work policy

In both the US and UK, performance measures and incentives have been used as a tool to drive up the effectiveness of services aiming to place disadvantaged people in work. Since the client group tends to be disadvantaged, agents who deliver services may face incentives to cream-skim their clients or otherwise manipulate the system; evidence for this has been found in American employment programmes (Anderson et al., 1993; Courty & Marschke, 1997; Heckman et al., 1997). Clearly some thought should be given to mechanisms that might help to prevent this occurring. One example, noted by (Cragg, 1997), might be to measure value-added outcomes rather than gross ones so as to take into account the outcome-relevant characteristics of the area or the client group. This would also reduce the risk of penalising service providers in disadvantaged areas.

UK-specific evidence in this area comes from Burgess et al. (2004), who analyse the performance Jobcentre Plus staff in response to a piloted incentive scheme. Their results lend some support to the free-rider problem that can arise with group-based incentives: they found that the impact on performance was higher in smaller teams (where staff can monitor each other's productivity) and smaller in larger teams (where it is easier to free-ride on the efforts of others). Thus the size of the groups involved in delivering a service is an important consideration. The researchers found that the incentive scheme used in this setting did not lead to any significant improvement in quality measures, only quantity ones – another conjectured side-effect. Nevertheless, they conclude that the scheme is more cost-effective than a general pay rise.

Winsor et al. (2010) outline five key issues that require thought when designing an incentive scheme in this area, but these criteria carry over well to the environment of early-years services. The issues are: (i) whether to provide individual-level or group-level bonuses (given that workers may be organised into teams); (ii) the size of the bonuses (or penalties) involved; (iii) the choice of performance measure; (iv) whether to reward all agents that reach a fixed target, or reward the best-performing agents only; (v) whether to place more weight upon improvements delivered to disadvantaged or hard-to-reach client groups.

4.3. Insights from health policy

Interestingly, research examining the application outcome-contingent funding or compensation in healthcare settings has produced generally negative conclusions. However, this may be because the healthcare sector is not particularly suited to such a framework or because such a framework has not been introduced in an appropriate manner. Rosenthal & Frank (2006) argue that paying for performance in healthcare has been driven by conceptual or ideological concerns rather than empirical ones, and may in fact have resulted in some cream-skimming. They are not the only researchers to draw out unintended consequences: Croxson et al. (2001) and Smith and York (2004), who look at the GP fundholder scheme and GP quality assessments respectively, find that anticipation effects were prevalent which resulted in manipulation of the systems in question.

Perverse incentives notwithstanding, the evidence is mixed on the efficacy of incentives in this environment. Town et al. (2005) review the relevant literature and conclude that the quality of healthcare delivery is not responsive to financial rewards, although they conclude that this may be because the rewards studied were small in magnitude. On the other hand, a trial conducted by Lindenauer et al. (2007) found that the combination of paying hospitals for performance with compelling them to report outcomes publicly, did lead to an improvement in healthcare. This suggests that financial incentives may need to be combined with other accountability devices in order to be effective. The desirability of a broad domain of measures and mechanisms is endorsed by the review in Conrad & Perry (2009). They suggest that superior clinical outcomes could be driven by striking an appropriate balance across a range of areas: specifically, in regard to bonuses or penalties, group-level or individual-level incentives, the size of the incentive, and the use evidence-based findings suitably modified to suit the specific circumstances of the sector.

4.4. High-level recommendations

This leads on to the most general lessons that can be learned – from both theory and application – which are most likely to be transferable to an early-years setting.

Ingraham (1993) examines the use and effectiveness of performance-related-pay in the public sector. While this may not seem highly relevant, some of the concluding messages can be transported. For example, it is advised to take public sector organisations and their needs as a first point of departure, rather than the rationale or experiences of payment for results in the private sector. This is similar to the recommendation in Burgess & Ratto (2003) that incentive schemes should be tailored to suit the circumstances of the specific agency in question.

Another recommendation by Ingraham (1993) is to dramatically increase the provision of data to planners and authorities: not only does this improve managerial information and planning, but it also serves as an accountability tool. Propper and Wilson (2003) concur with this and stress the need to build up an evidence base first before any large-scale implementation of payment by results or other performance management. They claim there to be a significant lack of evidence on the effectiveness of performance management mechanisms; this is not a reason not to implement them, however. Instead, the implementation of such measures must be via small-scale pilots that can be robustly evaluated to produce empirical evidence at the same time. The way that data are gathered also requires attention: independent data sources (such as official statistics or surveys of the experiences of client groups), which cannot be manipulated by agencies, should be used as much as possible. As stated above, it is preferable to measure value-added outcomes relative to some pre-policy benchmark. Furthermore, if data quality is an issue and it is not possible to measure outcomes precisely then incentives should be made less sharp, to stabilise the behavioural response of agencies and subsequent payments received (Burgess & Ratto, 2003).

The final high-level recommendation from this literature taken as a whole would be to examine the nature of the services being delivered and use such insights to tailor performance improvement schemes appropriately. A common feature of public service delivery, and a complication of the standard theoretical model, is that agencies and workers may have multiple competing tasks, while output itself may be difficult to measure. In such circumstances, direct supervision of agencies combined with subjective performance assessments may be more effective than payments or sanctions on the basis of objective indicators (Burgess & Metcalfe, 1999). This is also true if the service commissioner has more information about the process required to achieve outcomes, than about the quantity or quality of those outcomes (Savedoff, 2010). Paying for performance is therefore more suited to structures of service delivery where commissioners may have less of an idea of how agencies should achieve the required outcomes – which will be left to their discretion – but are able to measure the outcomes of service delivery well.

5. Conclusions

This document has summarised the rationale for payment by results, the possible issues arising from cognitive biases in behavioural economics, and the lessons to be learned from previous experiences of financial incentives in public service delivery. The overall message from all the

evidence together is that a measured approach seems the most preferable and the most likely to succeed. Microeconomic, private-sector theories have an intuitive appeal to them, but they may have limited applicability to public service delivery. With that in mind, incentive schemes should be tailored to reflect the realities of the specific organisations and processes involved in the delivery of early years services. The human aspect also needs to be taken into account, since the workers in question may differ from the *homo economicus* upon which textbook models of financial incentives are predicated. This requires very careful design of an incentive scheme, not just to minimise unintended consequences due to cognitive biases, but also to ensure that staff and agency morale is protected. Negative perceptions or reactions to an incentive scheme could potentially undermine any productivity gains encouraged by it. Finally, the implementation of such a scheme should be preceded, accompanied and followed by rigorous collection of independent data. This not only facilitates payment by results in the first place; it also enables its efficacy to be assessed, and can provide valuable evidence to support or direct future policy formation.

References

Anderson, K. H., Burkhauser, R. V. & Raymond, J. E. (1993), "The Effect of Creaming on Placement Rates Under the Job Training Partnership Act", Industrial and Labor Relations Review, Vol. 46, No. 4, pp. 613-624.

Burgess, S. & Metcalfe, P. (1999), "Incentives in organisations: A Selective Overview of the Literature with Application to the Public Sector", CMPO Working Paper 99/016.

Burgess, S., Propper, C., Ratto, M. & Tominey, E. (2004), "Incentives in the Public Sector: Evidence from a Government Agency", CMPO Working Paper No. 04/103.

Burgess, S. & Ratto, M. (2003), "The Role of Incentives in the Public Sector: Issues and Evidence", *Oxford Review of Economic Policy*, Vol. 19, No. 2, pp. 285–300.

Conrad, D. A. & Perry, L. (2009), "Quality-Based Financial Incentives in Health Care: Can We Improve Quality by Paying for It?", *Annual Review of Public Health*, Vol. 30, pp. 357–371.

Courty, P. & Marschke, G. (1997), "Measuring Government Performance: Lessons from a Federal Job-Training Program", *American Economic Review*, Vol. 87, No. 2, pp. 383-388.

Cragg, M. (1997), "Performance Incentives in the Public Sector: Evidence From the Job Training Partnership Act", *Journal of Law, Economics, & Organization*, Vol. 13, No. 1, pp. 147-168.

Croxson, B., Propper, C. & Perkins, A. (2001), "Do doctors respond to financial incentives? UK family doctors and the GP fundholder scheme", *Journal of Public Economics*, Vol. 79, pp. 375–398.

Dixit, A. (2002), "Incentives and Organizations in the Public Sector: An Interpretative Review", *Journal of Human Resources*, Vol. 37, No. 4, pp. 696–727.

Eldridge, C. & Palmer, N. (2009), "Performance-based payment: some reflections on the discourse, evidence and unanswered questions", *Health Policy and Planning*, Vol. 24, No. 3, pp. 160–166.

Gerhart, B., Trevor, C. & Graham, M. (1996), "New directions in employee compensation research", in *Research in Personnel and Human Resources Management*, ed. Ferris, G.R., pp. 143–203. Greenwich, CT: JAI Press.

Gneezy, U. & Rustichini, A. (2000), "Pay Enough or Don't Pay at All", *Quarterly Journal of Economics*, Vol. 115, No. 3, pp. 791–810.

Heckman, J., Heinrich, C. & Smith, J. (1997), "Assessing the Performance of Performance Standards in Public Bureaucracies", *American Economic Review*, Vol. 87, No. 2, pp. 389–395.

Ingraham, P. W. (1993), "Of Pigs in Pokes and Policy Diffusion: Another Look at Pay-for-Performance", *Public Administration Review*, Vol. 53, No. 4, pp. 348-356.

Irlenbusch, B. & Sliwka, D. (2005), "Incentives, Decision Frames, and Motivation Crowding Out – An Experimental Investigation", IZA Discussion Paper No. 1758.

Levin, I. P., Schneider, S. L. & Gaeth, G. J. (1998), "All Frames Are Not Created Equal: A Typology and Critical Analysis of Framing Effects", *Organizational Behavior and Human Decision Processes*, Vol. 76, No. 2, pp. 149–188.

Lindenauer, P. K., Remus, D., Roman, S., Rothberg, M. B., Benjamin, E. M., Ma, A., & Bratzler, D. W. (2007), "Public Reporting and Pay for Performance in Hospital Quality Improvement", *New England Journal of Medicine*, Vol. 356, No. 5, pp. 486-496.

Luft, J. (1994), "Bonus and penalty incentives contract choice by employees", *Journal of Accounting and Economics*, Vol. 18, pp. 181–206.

Martin, G. (1994), "Performance-related Pay in Nursing: Theory, Practice and Prospect", *Health Manpower Management*, Vol. 20, No. 5, pp.10–17.

Milgrom, P. & Roberts, J. (1992), *Economics, Organization and Management*. Englewood Cliffs, NJ: Prentice-Hall.

Propper, C. & Wilson, D. (2003), "The Use and Usefulness of Performance Measures in the Public Sector", *Oxford Review of Economic Policy*, Vol. 19, No. 2, pp. 250–267.

Rosenthal, M. B. & Frank, R. G. (2006), "What Is the Empirical Basis for Paying for Quality in Health Care?", *Medical Care Research and Review*, Vol. 63, No. 2, pp. 135-157.

Ross, S. A. (1973), "The Economic Theory of Agency: The Principal's Problem", *American Economic Review*, Vol. 63, No. 2, pp. 134–139.

Rynes, S. L., Gerhart, B. & Parks, L. (2005), "Personnel Psychology: Performance Evaluation and Pay for Performance", *Annual Review of Psychology*, Vol. 56, pp. 571–600.

Savedoff, W. D. (2010), "Basic Economics of Results-Based Financing in Health", Results-Based Financing for Health.

Smith, P. (1990), "The Use of Performance Indicators in the Public Sector", *Journal of the Royal Statistical Society*, Vol. 153, No. 1, pp. 53–72.

Smith, P. & York, N. (2004), "Quality Incentives: The Case Of U.K. General Practitioners", *Health Affairs*, Vol. 23, No.3, pp. 112-118.

Town, R., Kane, R., Johnson, P. & Butler, M. (2005) "Economic Incentives and Physicians' Delivery of Preventive Care", *American Journal of Preventive Medicine*, Vol. 28, No. 2, pp. 234-240.

Tversky, A. & Kahneman, D. (1981), "The Framing of Decisions and the Psychology of Choice", *Science*, Vol. 211, No. 4481, pp. 453–458.

Tversky, A. & Kahneman, D. (1986), "Rational Choice and the Framing of Decisions", *Journal of Business*, Vol. 59, No. 4, pp. S251–S278.

Van de Weghe, P. & Bruggeman, W. (2006), "The Impact of the Number of Performance Measures and Incentive Framing on Performance in a Multidimensional Task Environment", Working Paper 2006/404, Universiteit Gent.

Van Thiel, S. & Leeuw, F. L. (2002), "The Performance Paradox in the Public Sector", *Public Performance & Management Review*, Vol. 25, No. 3, pp. 267-281.

Weibel , A., Rost. K & Osterloh, M. (2010), "Pay for Performance in the Public Sector—Benefits and (Hidden) Costs", *Journal of Public Administration Research and Theory*, Vol. 20, No. 2, pp. 387-412.