

Evaluating and Rewarding the Quality of Teachers

INTERNATIONAL PRACTICES



ORGANISATION FOR ECONOMIC CO-OPERATION AND DEVELOPMENT

The OECD is a unique forum where the governments of 30 democracies work together to address the economic, social and environmental challenges of globalisation. The OECD is also at the forefront of efforts to understand and to help governments respond to new developments and concerns, such as corporate governance, the information economy and the challenges of an ageing population. The Organisation provides a setting where governments can compare policy experiences, seek answers to common problems, identify good practice and work to co-ordinate domestic and international policies.

The OECD member countries are: Australia, Austria, Belgium, Canada, the Czech Republic, Denmark, Finland, France, Germany, Greece, Hungary, Iceland, Ireland, Italy, Japan, Korea, Luxembourg, Mexico, the Netherlands, New Zealand, Norway, Poland, Portugal, the Slovak Republic, Spain, Sweden, Switzerland, Turkey, the United Kingdom and the United States. The Commission of the European Communities takes part in the work of the OECD.

OECD Publishing disseminates widely the results of the Organisation's statistics gathering and research on economic, social and environmental issues, as well as the conventions, guidelines and standards agreed by its members.

This work is published on the responsibility of the Secretary-General of the OECD. The opinions expressed and arguments employed herein do not necessarily reflect the official views of the Organisation or of the governments of its member countries.

This publication is a product of the co-operation agreement established between the Government of Mexico and the OECD, which aims to improve the quality of education in Mexico.

ISBN 978-92-64-06198-9 (print)
ISBN 978-92-64-03435-8 (PDF)

Photo credits: ©UNESCO/José Gabriel Ruiz Lembo.

Corrigenda to OECD publications may be found on line at: www.oecd.org/publishing/corrigenda.

© OECD 2009

You can copy, download or print OECD content for your own use, and you can include excerpts from OECD publications, databases and multimedia products in your own documents, presentations, blogs, websites and teaching materials, provided that suitable acknowledgment of OECD as source and copyright owner is given. All requests for public or commercial use and translation rights should be submitted to rights@oecd.org. Requests for permission to photocopy portions of this material for public or commercial use shall be addressed directly to the Copyright Clearance Center (CCC) at info@copyright.com or the Centre français d'exploitation du droit de copie (CFC) at contact@cfcopies.com.

CHAPTER 3

Design Components of Incentive Pay Programmes in the Education Sector

| | |
|---|----|
| Introduction | 66 |
| Incentive structure | 67 |
| • Rank-order tournaments..... | 67 |
| • Fixed performance contract..... | 67 |
| Unit of accountability | 71 |
| • Individual unit of accountability..... | 71 |
| • Group unit of accountability..... | 72 |
| • Hybrid models of accountability..... | 74 |
| Performance measures | 75 |
| • Educational inputs..... | 75 |
| • Educational outputs..... | 79 |
| • Multiple measures – linking educational inputs and outputs..... | 82 |
| Performance standards and thresholds | 83 |
| • Linear models..... | 83 |
| • Step functions..... | 83 |
| • Limited linear models..... | 84 |
| • Threshold levels..... | 85 |
| Size and distribution of bonus awards | 85 |
| • Size of bonus..... | 85 |
| • Distribution of award..... | 87 |
| • Non-monetary incentives..... | 88 |
| Payout frequency | 89 |
| Cultural considerations | 90 |
| Conclusion | 91 |
| Notes | 92 |
| References | 94 |

INTRODUCTION

Education reform is increasingly focused on human capital formation because high-quality teachers are seen as the most direct and effective pathways to improving student achievement. Based on the argument that prevailing compensation practices provide weak incentives for teachers, and that inefficiencies arise from rigidities in current compensation policies, several national systems of public education have explored teacher compensation reforms (Podgursky and Springer, 2007).¹ Proponents argue that financial incentives can motivate teachers to achieve higher levels of performance, entice more effective teachers to join or remain in the teaching profession, and align teacher behaviours and interests with institutional goals.

In spite of the intuitive appeal incentive pay has for some stakeholders, a sturdy and influential base of individuals and organisations fundamentally oppose its use in education (Eberts, 2007; Goldhaber, 2009; Kingdon and Teal, 2008). Opponents contend that such pay renders schools less effective by crowding out intrinsic rewards; that is, teachers will lose interest in the profession as they are increasingly rewarded for student outcomes. Critics further argue that the education system lacks appropriate measures for evaluating teacher performance, that reward programs will decrease teacher collaboration, and that the current body of evidence on the impact of pay-for-performance programmes is inconclusive.

Recent experimental and quasi-experimental evidence paints a mixed picture of the impact of teacher incentive-pay programmes. Muralidharan and Sundararaman (2008) and Lavy (2002, 2007) found that teacher incentive programs in **India** and **Israel**, respectively, improved student outcomes and promoted positive changes in teacher behaviour and/or classroom pedagogy. Glewwe, Ilias and Kremer (2008) similarly reported that students instructed by teachers eligible to receive a bonus award in **Kenya** demonstrated better scores on high-stakes tests; however, no discernible impact was found on low-stakes tests taken by treatment group students or on the same students when they took high-stakes tests during the post-intervention school year.

The design components of incentive programmes in the education sector have received little attention in the literature, particularly with regard to the relative advantages and limitations of different options for measuring and rewarding teachers. Since the design of an incentive programme can lead to dramatic differences in its impact on students, teachers, and administrators, it is important to learn more about how specific design elements affect teachers' responses.

The following discussion is largely informed by the theoretical literature on incentive systems as well as empirical evidence from evaluations of past and present programmes and policies in the **United States (U.S.)** and other countries. We draw on experiences from both within and outside the education sector to assess specific design components of incentive-pay programmes and policies and their potential impact on student outcomes, teacher attitudes and behaviour, and institutional dynamics. Specifically, we assess:

- incentive structure;
- unit of accountability;
- performance measures;
- performance standards and thresholds;
- size and distribution of bonus awards;
- payout frequency; and
- cultural considerations.

This chapter presents a brief discussion of the relationship between incentive-pay programme characteristics and cultural context before offering concluding remarks.

INCENTIVE STRUCTURE

Incentive structure refers to the scheme or mechanism that guides the allocation of awards in a pay-for-performance system. In some cases only a limited number of employees can earn an award, while in others any employee who meets a predetermined performance standard will receive an award. The two main forms of incentive structure are rank-order tournaments and fixed performance contracts, each of which, under certain conditions, could be an optimal labour contract yielding the first best outcome. Both forms represent very different compensation schemes with distinct advantages and limitations.

Rank-order tournaments

Rank-order tournaments are incentive structures that limit performance-based rewards to a fixed percentage of the relevant work group. The key feature that distinguishes tournaments from other incentive-pay structures is that compensation depends on relative performance rather than absolute performance. For example, an individual level pay-for-performance experiment implemented in **Israel** and evaluated by Lavy (2004) compared the performance of teachers in similar grades and subjects and then awarded individual bonuses to the teachers with positive gains scoring in the top performance quartiles.

A primary strength of tournament incentive structures is lower information costs (Baker, Jensen and Murphy, 1988). For example, if a school district announced that the top ten teachers in a school could win bonuses of USD 5 000, with 20 schools in the district participating in the programme, and each school had 40 teachers, the school district could anticipate that the cost of the bonuses would be USD 1 million a year (10 winning teachers x 20 schools x USD 5 000 per winning teacher). However, if the incentive structure is designed so that all teachers could potentially win an award by exceeding a fixed benchmark, the financial exposure is less certain and potentially much greater. The cost of operating the programme could range from USD 0 (no teachers meet the standard) to USD 4 million (40 winning teachers x 20 schools x USD 5 000 per winning teacher).

Stakeholders wanting to integrate a rank-ordered tournament as part of an incentive-pay programme should explore constructing comparable peer groups to measure and evaluate the performance of individual teachers or teams of teachers. Neal (2009) suggests that school systems construct league-specific tournaments; that is, a school's performance is considered relative to the performance of schools that serve similar types of students and families. Without these comparison sets of homogenous schools that are regularly updated, Neal argued, it is impossible to obtain credible estimates of a school's or teacher's contribution to student achievement outcomes. Conceivably, this notion of league-specific tournaments can be applied at the classroom-level while further specifying the matching criteria on which leagues are formed (*i.e.* subject taught, number of classes taught, average class size), thus minimising within-school variation.

Rank-order tournaments can suffer from one well-recognised defect: promotion of competition among teachers or groups of teachers can lead to a breakdown in the collegiate ethos, thereby reducing rather than increasing performance. This can become a serious threat if teachers in the same school are no longer willing to help one another. For example, the grade five mathematics teacher may opt against sharing a lesson plan with the other grade five mathematics teacher because they are competing for the same pot of money. Indeed, allowing only some teachers, rather than all who met a predetermined standard, to receive a bonus was attributed with the collapse of many merit-pay programmes in the 1980s.

Making high-stakes comparisons of teachers or teams of teachers within a school can encourage teachers to sabotage the performance of their peers (Burgess *et al.*, 2001; Neal, 2009). Murnane and Cohen (1985) suggested sabotage can take place in the form of teachers using friendships with students' parents to spread damaging rumours about a colleague's ability as a teacher. Gonring, Teske and Jupp (2007) did not find any "insider-evidence" of unhealthy competition, however, created by Denver's Professional Compensation System for Teachers (ProComp).

Another prime consideration in developing an incentive-pay programme is that district officials, legislators and other individuals responsible for funding bonuses tend to be risk-averse. Elmore, Abelman and Fuhrman (1996) noted that, “Although legislators and governors frequently ‘talk tough,’ they are very susceptible to pressures by school people to back down and soften policies perceived to be restrictive or punitive” (pp. 85-86). In their recent review of the New Public Management reform movement, which is inextricably linked to the resurgence of interest in performance-related pay policies, Perry, Engbers and Yun (2009, p. 14) similarly observed that, “It is ironic that the same politicians who promote performance-related pay also may vote against appropriations to fund it if they perceive fiscal restraint serves larger political ends.”

Incentive-pay reforms are also vulnerable to shifts in political leadership and public opinion. When supporting legislators, superintendents and/or school board members leave office, the political will to continue educator incentive-pay programmes can disappear rapidly (Ballou, 2001; Ballou and Podgursky, 1997; McCollum, 2001). Given the uncertain and potentially substantial costs, incentive-pay programmes based on fixed performance contracts, for example, are likely to be among the first line items on the political chopping block. Thus it is important to construct an incentive-pay programme that appears financially practical to a wide range of educators, as well as to elected and appointed officials.

Fixed performance contract

The second major form of incentive structure is the fixed performance contract, which defines the performance standard teachers, teams or schools must meet to earn an award. The standard can take on a number of forms, including a single threshold, multiple thresholds (e.g. a step function), or a continuous, linear standard after some threshold has been met. In contrast to a rank-order tournament, any teacher who meets the predetermined performance standard benefits regardless of the performance of other teachers.

The primary rationale for adopting a fixed performance contract is pragmatic – fixed performance contracts avoid competition among teachers. Teaching is characterised by widespread compliance to an implicit or explicit code of professional ethics which can be driven by self-interested reputation, internalised morality, peer esteem and a service ideal (Mallick, unknown). Competition may threaten professional norms, particularly those related to collaboration and teamwork, which some studies have found to be key ingredients in highly effective schools (Bryk and Schneider, 2002; Hallinger and Murphy, 1986).

Many teachers’ unions strongly oppose the introduction of competitive pressures into schools or school systems, irrespective of whether pressures come in the form of an incentive-pay policy or from market competition, such as school choice. In an analysis of multiple-waves of data collected in the U.S. Department of Education’s Schools and Staffing Survey, Ballou (2001) found a strong inverse relationship between the use of pay-for-performance programmes and the degree of union influence. Tomlinson (2000) similarly reported that the National Union of Teachers opposed pay-for-performance proposal in England and Wales because:

The Government’s proposals would profoundly damage the professional culture and teamwork that is at the heart of successfully managed schools. The pressures caused by the performance management structure would generate distrust and counter-productive competition between colleagues. Teachers could be demotivated and deprived of the confidence they need to be successful in the classroom. Stress, anxiety, apprehension and self doubt are not conducive to successful teaching (p. 6).²

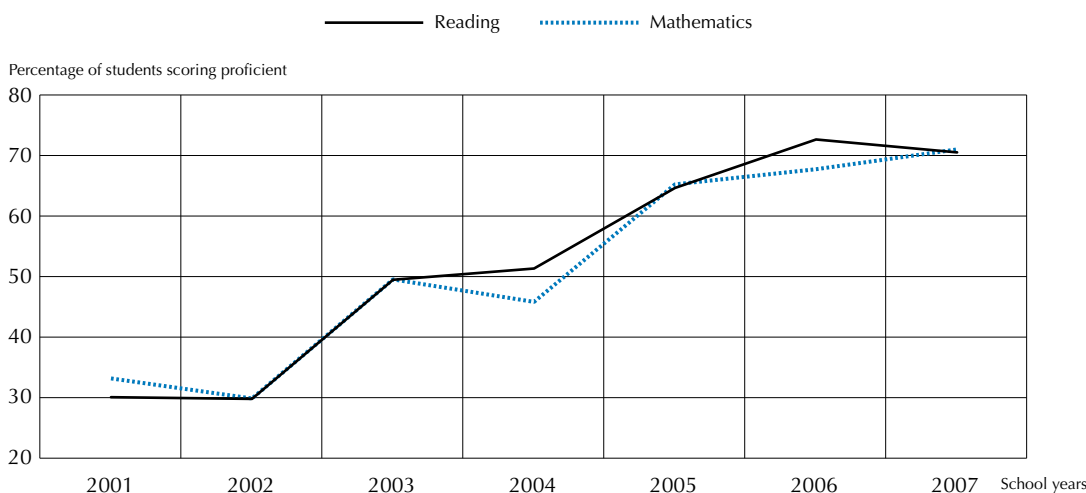
Fixed performance contracts set clear standards which enable teachers to set personal goals in relation to those standards. Goal-setting theory, which refers to the establishment of a purpose in reference to a performance standard, helps direct attention and action, further motivating individuals toward the standard

and, ultimately, leading to better task performance (Locke, 1968; Locke, Latham and Smith, 1990; Locke, Shaw, Saari and Latham, 1981). Recognizing that rank-order tournament incentive schemes are contingent on a “do-your-best” mentality, teachers are more likely to respond to a fixed performance standard because a relative performance standard can be perceived as impossible to attain, or teachers may conclude the evaluation system is biased because teachers do not know what standard they need to achieve in order to earn an award until after the fact.

Conversely, the large financial exposure inherent in most performance contracts has been associated with teachers’ skepticism about incentive-pay policies. In the **United Kingdom’s (U.K.)** Performance Threshold and Upper-Scale Pay System, teachers and principals questioned whether adequate funding existed to implement and sustain the programme given there was no quota on the number of teachers who could earn an award (Adnett, 2003; Cutler and Waine, 2004; Marsden and Belfield, 2006). Skepticism can reduce the likelihood that teachers will increase their effort in pursuit of a bonus award. It can also increase opportunistic behaviour on the part of teachers, and can decrease their level of buy-in (Hamilton, 2005; Kelley, 2002; Lewis and Springer, 2008).

Figure 3.1

Percentage of schools with a minimum of 72% of third grade students scoring proficient in reading and mathematics



Source: OECD.

Fixing in advance the total number of individuals, teams of teachers, or schools that can earn bonuses (*i.e.* a rank-ordered tournament) helps mitigate several complicating factors associated with judging their performance. For example, Figure 3.1 displays the percentage of schools with at least 72% of their third-grade students scoring proficient in reading and mathematics as measured by **Minnesota’s** minimum competency accountability programme. The percentage of schools meeting the 72% standard more than doubled from the 2000-01 to 2006-07 school year, an impressive increase particularly given the relatively short time series.

Now imagine the Minnesota legislature designed and implemented a performance-pay programme that rewards any performance above a minimum threshold, wherein the unit of accountability is the school. Assume the programme also relies on a single performance standard for determining bonus award eligibility – any school with at least 72% of its students scoring proficient receives USD 2 500 per full-time equivalent instructional employee. In the first year of the programme (2001-02 school year), this would have meant that approximately 255 schools, or roughly 30% of eligible schools, met the performance standard, translating to USD 19.2 million in bonus awards (we will assume 30 teachers per school throughout this hypothetical example).

Since the legislature opted for a fixed performance contract incentive structure, the amount of resources necessary to remunerate teachers increases dramatically in a relatively short period of time. In the 2002-03 school year, more than 420 schools met the performance standard, which means bonus awards totaled around USD 32.8 million according to the hypothetical parameters. Two years later, more than 560 schools met the standard, adding an additional USD 10 million to the 2002-03 figure. This would place a huge financial burden on the taxpayer, as well as on the budget of the agency operating the programme.³

Some literature suggests school personnel and bargaining organisations are prone to exploit the fixed performance contract incentive system. Employees “expend time, effort, and ingenuity” to influence others’ decision-making in an attempt to maximise their own benefit (Holmstrom and Milgrom, 1987). Since the design (or reauthorization) of an incentive-pay programme will involve teachers, and/or the bargaining unit protecting the interests of the “median” teacher, these individuals may advocate for a lower performance standard (Indjejikian, Lenk and Nanda, 2000; Merchant and Manzoni, 1989; Murphy, 1999). As Heneman, Milanowski and Kimball (2007) noted:

Over time, teachers exert pressure to lower performance standards, increasing the number of teachers who become eligible for performance pay... while such an eventuality may enhance acceptance of the plan, it also drives up costs, creating long-term funding issues (p.6).

School systems also need to be mindful of a potential ratchet effect when elevating the performance standard (Freixas, Guesnerie and Tirole, 1985; Weitzman, 1980). If a school system increases the rigour of the performance standard that an individual has to meet to earn a bonus award, teachers and other school personnel may perceive the heightened award threshold as a punishment. If so, the recalibration process can potentially reduce productivity, compromise trust and information sharing among school administrators, schools, and teachers, and create sufficient opposition such that the future of the incentive-pay programme is put into question (Lazear, 1995). One strategy to overcome negative reaction to increasing the standard is to increase the size of the bonus or the percentage of employees eligible to earn the bonus at the same time the standards are revised.

Another consideration in determining performance standards is ensuring the standard is not biased in favour of particular schools, teams, or teachers. In their evaluation of **New York City’s** School-Wide Performance Bonus Program (SPBP), Springer and Winters (2009) studied the relationship between the performance target assigned to particular categories of schools and the probability that schools in those categories met at least part of their performance target. Schools were classified according to the previous year’s progress-report data, with Category One schools being the highest performing and Category Five the lowest. According to the authors, the probability of Category Four schools and Category Five schools earning at least part of their performance bonus award was nearly ten times greater than the probability of Category Three and Category Two schools. While the authors’ emphasised these findings came from the first year of the programme, if a similar pattern emerges in subsequent years one might argue the SPBP is rewarding mediocrity given that a percentage of schools in Category Four and Category Five will meet their performance target irrespective of any effort put forward. Moreover, these schools were the lowest performing schools identified by the district’s accountability system.

UNIT OF ACCOUNTABILITY

The next design component of an incentive-pay programme is the unit of accountability. The unit of accountability refers to the entity responsible for a measurable product or service whose performance on that measurable dimension determines bonus eligibility. The unit of accountability can be defined in a myriad of ways, including the individual teacher, a grade-level or departmental team of teachers, all employees within a school, or some combination thereof.

There are distinct advantages and limitations that policy makers need to carefully consider when defining the unit of accountability in incentive-pay programmes. This next section begins with a discussion of the individual as the unit of accountability and then shifts attention to the group, where the group may be defined as a within-school team of teachers or an entire school taken as a single unit. The final section provides a similar review of hybrid models of accountability.

Individual unit of accountability

In an incentive-pay plan that relies on an individual unit of accountability, the performance of the individual teacher determines award eligibility. The individual unit of accountability creates the strongest connection between variation in award size received by teachers and the variation in teacher effectiveness. Since the inability to directly impact one's chances of earning a performance award can lead to decreased effort, the individual unit of accountability is often identified as optimal because the individual assumes sole responsibility for his or her performance (Freeman and Gelber, 2006).

Incentive-pay programmes in the education sector that reward teachers at the individual level of accountability maximise conditions for altering the composition of the teacher labour force. The most effective teachers in the system are more likely to be retained, sending a strong signal to those teachers not receiving a reward to improve or exit. Additionally, prospective teachers most likely to be rewarded under the performance appraisal system are more likely to join the profession. Thus, as Podgursky and Springer (2007) argue, teacher turnover would become part of a virtuous cycle of quality improvement, rather than a problem to be minimised.

The theoretical underpinnings of an incentive-pay programme that rewards individual teacher performance is critically important in light of inefficiencies in current teacher hiring practices and labour market selection (Lazear, 2000). In terms of hiring practices, principals and building administrators must use noisy signals of "true" teacher effectiveness, such as years of experience, highest degree held, or past-employer recommendations. Informational deficiencies in the hiring process are overcome in most professions by employee performance assessments and the close coupling of pay increases and promotion decisions with actual productivity. Without such structures in education, combined with an often insurmountable tenure system that makes contract non-renewal difficult, pay for performance becomes all the more relevant.

In terms of labour market selection, an incentive-pay programme tends to attract and retain individuals who are particularly good at the activity to which incentives are attached, while repelling those who are not. While incentives can raise the productivity of the typical worker, an incentive system can also raise the overall quality of the workforce simply through differential recruitment and retention of more effective workers (Podgursky and Springer, 2007). Lazear's (2000) widely-cited case study of Safelite Glass Corporation, for example, reported that transitioning the company's compensation system from hourly wages to piece rates was associated with a 44% increase in worker productivity, half of which resulted from the sorting of more able workers. In a case study of a **British Columbia** tree-planting firm, Paarsch and Shearer (2000) similarly found increases in productivity attributable to both increased effort and sorting. Of course, the production process and output for teachers and schools is much more complex.

A recent and provocative study of teacher turnover found evidence that the migration of high ability women out of teaching between 1960 and the present primarily resulted from the “push” of teacher pay compression, which took away relatively higher earnings opportunities for teachers, as opposed to the “pull” of more lucrative non-teaching opportunities (Hoxby and Leigh, 2004). Although remunerative opportunities outside of teaching for teachers of high and low ability grew over this period, Hoxby and Leigh (2004) argued it was pay compression within the education system that accelerated the exit of higher ability teachers. Similarly, Corcoran, Evans and Schwab (2004) found the likelihood that a female in the top of her upper-secondary school class will enter teaching decreased dramatically between 1957 and 1992.⁴

However, the literature on optimal incentive structures questions whether the individual as the unit of accountability is the optimal incentive design when outputs cannot easily be attributed to an individual teacher. If work tasks of two or more individuals influence the measured and rewarded activity in an incentive-pay programme, relying on the individual as the unit of accountability may cause some teachers to take exception to high performers because they may affect the standard upon which all other teachers are judged (Baron and Kreps, 1999; Bowles and Gintis, 2002; Deutsch, 1985). Moreover, as discussed previously, rewarding teachers for their individual performance is also contrary to the highly collaborative nature of teaching and may even reduce the incentive to cooperate with others (Mallick, date unknown; Milgrom and Roberts, 1990; Murnane and Cohen, 1986).

There are numerous other interdependencies that complicate efforts to isolate the contribution of an individual teacher to student outcomes. If the individual is the unit of accountability, how should the evaluation system take into account the cumulative effects of learning from one school year to the next? Are there practical strategies for taking into consideration depreciation or appreciation in learning over time, and whether the instructional strategies of the prior year's teacher influence learning gain or loss (*i.e.* teaching to the test versus teaching higher order skills)? Identifying the individual contribution of a teacher to student learning is further complicated, for example, when students enroll in a reading and an English language arts course, each of which is taught by a different instructor.⁵ Individual incentive-pay programmes can also create resentment among teachers, as teachers not receiving an award may believe they contributed to the success of students linked with a teacher that receives a bonus award.

Factors influencing student learning also extend outside the school building walls. For example, in their highly influential work on consumer preferences and school choice, Schneider, Teske and Marshall (2000) noted that interactions among students and parents make parents co-producers of educational outcomes; parents undoubtedly shape the educational goals of their children. Social science research has also identified a number of neighbourhood factors that can influence student outcomes, including social disorganisation, community resources, environmental deterioration (*e.g.* abandoned buildings), crime and delinquency (*e.g.* drug dealing and violent crime) (Ginther, Haveman and Wolfe, 2000; Leventhal and Brooks-Gunn, 2004; Wilson, 1987). These and other complicating factors make it difficult to isolate an individual teacher's contribution to student outcomes.

Group unit of accountability

The group unit of accountability refers to award eligibility as the product of aggregated performance among members of a group, where the size of a group can range from as few as two employees to all employees within a firm. Sometimes referred to as profit sharing, gain sharing, team incentives, goal sharing, achievement sharing, win sharing, and results sharing (Gaynor and Pauly, 1990; Holmstrom, 1982), group incentive structures have experienced tremendous growth in popularity. In the **U.S.**, the number of group incentive-pay programmes operating in firms increased from just over 2 000 in 1945 to nearly 500 000 in 1991 (Nalbatian and Schotter, 1997).

It is important to stress that the group unit of accountability is not limited to the school-level, a common mistake encountered among incentive-pay programmes throughout the education sector. Group incentive systems can take on a number of different forms, including monitoring and rewarding performance of grade-level teams of teachers, disciplinary or inter-disciplinary departments of teachers, or any other recognizable unit within the school.

Organisational theory suggests group incentives can promote social cohesion, feelings of fairness, and productivity norms (Lazear, 1998; Pfeffer, 1995; Rosen, 1986). Improved social cohesion among workers can foster knowledge transfer and mutual learning that result in increased productivity in the long run (Che and Yoo, 2001). For example, in the case study of garment plants reported in Berg *et al.* (1996) and Hamilton, Nickerson and Owan (2003), the formation of teams with workers of varying abilities facilitated interactions among high- and low-ability workers so that more able workers taught less effective workers how to better execute tasks and become more productive.

Although productivity gains from knowledge transfer and mutual learning may yield results similar to an individual incentive programme, some empirical research describes a much more nuanced story when firms have adopted team-based incentive programmes. Weiss (1987) and Hansen (1997) found that while the performance of lower-ability workers improved, the productivity among the highest producers decreased. Thus the group incentive-pay programme may not have the same compositional effect on the workforce and, as a consequence, the expected benefit of the group unit of accountability may be all for not.

The free-rider problem and shirking are other concerns designers of incentive-pay programmes need to recognise when considering the group as the unit of accountability. If the school is the unit of accountability, or the size of the within-school team is greater than six to eight teachers, specific individuals on a team may be less likely to shoulder their fair share of the workload. They know the capabilities of teammates can make up for their subpar performance. Because an individual who exerted minimal effort will still receive a bonus award if the performance of the other members of his team offset his mediocre effort, group incentive systems inevitably result in the inefficient allocation of some resources.

Kandel and Lazear (1992) and others have argued that as long as the size of a within-organisation team is not too large, the free-rider problem can be solved through peer pressure. For instance, peer monitoring and the enforcement of social penalties in the form of shame, guilt, empathy, and mutual monitoring can lead to individual team members being accountable for their performance to other members. If a teacher has both monetary and social incentives to not shirk, Kandel and Lazear (1992) contend the motivational forces that would have been “choked off” by the free-rider problem are recovered.

However, even though the theoretical literature has argued that the free-rider problem can be overcome in certain situations, a large-scale random assignment evaluation revealed individual teacher incentives elicited greater gains and productivity as compared to group incentive programmes. Known as the Andhra Pradesh Randomized Evaluation Study (AP RESt), the impact of two output-based incentive systems in **India** (an individual teacher incentive programme and a group-level teacher incentive programme) and two input-based resource interventions (one provided an extra-paraprofessional teacher and another provided block grants) were evaluated. Muralidharan and Sundararaman (2008) found that students enrolled in a class instructed by a teacher selected for the group incentive intervention outperformed students in control condition classes that received only their usual allocations of resources, on both the mathematics and language exams (0.28 and 0.16 standard deviations, respectively).⁶ Students enrolled in schools assigned to the individual incentive condition outperformed students in both the group incentive condition and the control condition after the second year of implementation. In terms of cost effectiveness, moreover, the authors reported both of the incentive-pay interventions were preferable as compared to the input-based resource interventions and the control condition.

An incentive-pay programme relying on the group unit of accountability may also fail to recognise some of the most effective teachers in a school or school system if those individuals are assigned to a team with lower-ability teachers. High-ability teachers who are not rewarded may leave the school system or profession as a consequence of being discouraged with a performance management system that does not recognise their individual performance. Rivkin, Hanushek and Kain (2005) further deduced school-based incentive systems may also be a highly-inefficient approach for rewarding performance given the presence of significant within-school variation in teacher effectiveness. Recognizing that two of the key goals of incentive-pay programmes are to improve the composition of the labour force and organisational productivity, it is critically important that stakeholders designing an incentive system closely monitor and evaluate the programme for any signs of perverse sorting.

A final consideration when adopting a group incentive plan is procedural. How should the composition of within-school teams be formed? The purest approach to assigning teachers to teams is a simple random lottery wherein each teacher has an equal chance of being assigned to a given group (Lazear and Gibbs, 2009). Assuming the number of individuals in the lottery is large enough in relation to the number of teams, randomly assigning teachers to teams means work ability will be balanced across the teams. Another strategy is for a school principal or administrator to assign individual teachers to teams. Most firms rely on supervisors to select the composition of teams, which makes the most sense if the pool of candidates being assigned to teams is new to the company (Lazear and Gibbs, 2009). Lazear and Gibbs (2009) further noted an alternate draw is typically identified as the optimal strategy if schools want to make teachers responsible for the make-up of the within-school teams.⁷

Hybrid models of accountability

Hybrid models offer a third type of accountability system. A hybrid model joins elements of both independent and interdependent work, meaning the incentive structure considers the unique contribution of an individual teacher while also supporting teamwork and collaboration among teachers. Hybrid models are more commonly found in the **U.S.** and, as discussed later in this chapter, may be related to the cultural context in which an incentive system operates.

Initiated in 1999 by the Milken Family Foundation, the Teacher Advancement Program (TAP) is a well-known incentive-pay programme drawing on a hybrid model of accountability. Bonus calculations incorporate standardised tests scores for both individual teacher performance and the performance of the entire school. While each school can ultimately decide the exact percentage to attribute to each unit of accountability, TAP guidelines recommend a breakdown of 20% school level and 30% individual level.⁸

Establishing an appropriate balance among design components can pose a number of challenges. If too little weight is placed on some aspect of the schooling process perceived by teachers and other stakeholders as important, then the influence of the incentive runs the risk of being diluted to the point that the pay programme becomes irrelevant. Such a scenario can also send a strong signal that a particular aspect of the schooling process is not valued, which may lead to discouragement and a breakdown in trust between teachers and administrators. For example, in their study of teacher perceptions to **Florida's** Better Educated Students and Teachers (BEST) Teaching Salary Career Ladder Program, Mullen and Slagle (2007) reported that teachers perceived as highly-flawed the way in which the BEST program conceptualised teaching.

A hybrid model of accountability necessitates that the technical requirements of both group and individual units of accountability be in place. Systems that currently have the capacity for linking records between individual teachers and students will not have significant further requirements if also adopting aggregate units of accountability. Conversely, those school systems starting with a group as the unit of accountability may not have

the data linkage necessary for measuring performance at the individual teacher level, or the technical capacity to manage an incentive-pay programme with a more disaggregated accountability unit. Current capacity and future costs of using a hybrid model could be significant and should be investigated beforehand.⁹

A lack of goal clarity is more likely to affect the motivational effect of an incentive-pay programme relying on a hybrid unit accountability. In a summary report of performance-related pay policies prepared for the U.K.'s National Union of Teachers, Richardson (1999, p. 20) noted, "...a multiplicity of goals is likely to cause problems, because it reduces goal clarity. Being under pressure to meet many goals, especially where some of them are ambiguous, makes it difficult for employees to focus their efforts properly." Recognising the inverse relationship between the number of design components incorporated into incentive-pay programmes and teacher goal clarity, it is important for hybrid models of accountability to clearly specify how the programme works.

Some challenges regarding goal clarity can be mitigated if the programme design team establishes a multi-prong strategy to enhance teachers' and school personnel's understanding of the programme. Stakeholders can train school administrators on the ins-and-outs of the programmes to help facilitate implementation as well as buy-in among teachers. In fact, in their evaluation of school-based performance award programmes in **Kentucky** and **North Carolina**, Kelley (1999) reported programme buy-in is largely conditional on the confidence teachers have about the system functioning as intended and on the responsiveness of administrators to questions from teachers about the programme.¹⁰ Chapter 7 presents a detailed review of implementation and technical assistance issues.

PERFORMANCE MEASURES

Performance measures refer to the evaluation criteria for gauging employee performance. Despite many education systems having a long history of evaluating teachers and schools, there remains considerable variation in stakeholders' perceptions of what should be evaluated, how appraisal criteria should be linked to rewards, and which measures and instruments can reliably and accurately reflect performance.

Table 3.1 displays the primary performance measures implemented in a handful of high-profile incentive-pay programmes. Slightly more than half of the programmes include an input- or a process-oriented performance measure, which can include professional development, job enlargement, student attendance, and teacher attendance.¹¹ Educational outputs cover student achievement levels and gains at the student, teacher, and school-level as well as accumulation of credit hours and student retention rates.

Several programmes also incorporate recruitment and retention incentives or incentives for teaching in a hard-to-staff school or subject. Hard-to-staff school incentives are most often offered for teaching in high-needs schools or districts, typically either high-poverty, low-performing, or geographically remote schools. Like hard-to-staff subject incentives, these incentives are designed specifically to address market factors.¹² As presented in Table 3.1, we classify these market-based compensation reforms under educational inputs and processes.

Educational inputs

Educational inputs refer to measures that reward teachers for activities believed to improve instructional quality. Examples of the types of activities rewarded under an input-focused incentive system include completing an advanced degree, serving as a mentor or master teacher, taking on a leadership role or other special duties, participating in professional development activities, and teacher attendance. We first review more prominent types of input-focused programmes and then discuss perspectives regarding subjective measures of performance.

Table 3.1

Characteristics of incentive pay programmes in the United States and other countries

| | Period | Teacher Performance Measures | Size of bonus (USD, normal) | | |
|--|----------------|---------------------------------------|-----------------------------|-----------|------------------|
| | | Unit of Accountability | Minimum | Maximum | % Monthly Salary |
| United States | | | | | |
| School Incentive Program (Dallas, Texas) | 1992 - 1995 | School | \$450 | \$1 000 | ≈ 10% - 22% |
| Merit Pay Program (State of Michigan) | 1996 - 1997 | Teacher | \$1 000 | ≈ \$5,000 | ≈ 12.5% - 62.5% |
| Teacher Advancement Program (United States) | 1999 - present | Hybrid (Teacher and School) | \$2 500 | \$12 000 | ≈ 45% - 216% |
| Achievement Challenge Pilot Project (Little Rock, Arkansas) | 2005 - 2007 | Teacher | \$350 | \$7 600 | ≈ 8% - 174% |
| Governor Educator Excellence Grant Program (State of Texas) | 2005 - 2008 | Varies (Teacher, Team, and/or School) | \$75 | \$15 000 | ≈ 1.3% - 270% |
| Professional Compensation System for Teachers (Denver, Colorado) | 2005 - present | Hybrid (Teacher and School) | \$400 | \$2 500 | ≈ 7.2% - 45% |
| ASPIRE Program (Houston, Texas) | 2005 - present | Hybrid (Teacher and School) | \$250 | \$8 600 | ≈ 4.5% - 154.8% |
| POINT Experiment (Nashville, Tennessee) | 2006 - 2009 | Individual | \$5 000 | \$15 000 | ≈ 90% - 270% |
| Texas Educator Excellence Grant Program (State of Texas) | 2007 - present | Varies (Teacher, Team, and/or School) | \$20 | \$20 462 | ≈ 0.4% - 365% |
| Team Incentive Project (Round Rock, Texas) | 2008 - present | Team | ≈ \$6 000 | | ≈ 108% |
| School-Wide Performance Bonus Program (New York City, New York) | 2008 - present | School | \$1 400 | \$3 600 | ≈ 25.2% - 64.8% |
| International | | | | | |
| School Performance Program (Israel) | 1996 - 1997 | School | \$1 000 | \$2 500 | ≈ 30% - 75% |
| International Christelijk Steunfonds Incentive Program (Buso and Teso, Kenya) | 1998 - 1999 | School | \$26 | \$51 | ≈ 21% - 43% |
| Salario al Merito (Bolivia) | 1998 - 1999 | Teacher | \$300 | \$333 | ≈ 66% |
| Teacher-Incentive Experiment (Israel) | 2001 | Teacher | \$1 750 | \$15 000 | ≈ 10% - 40% |
| Carrera Magisterial (Mexico) | 1993 - present | Teacher | ... | ... | ≈ 27% - 224% |
| Sistema Nacional de Evaluación de Desempeño de los Establecimientos Educativos (Chile) | 1996 - present | School | \$439 | \$439 | ≈ 4.7% - 7.2% |
| Performance Threshold and Upper Pay Scale System (England) | 1999 - present | Individual | \$2 000 | | ≈ 9% |
| Incentivo Colectivo a Escuelas (Bolivia) | 2001 - present | School | \$281 | \$281 | ≈ 5% - 19% |
| Randomized Evaluation Project (Andhra Pradesh, India) | 2006 - present | Teacher or School | \$2.25 | \$450 | ≈ 33% - 50% |
| Performance Pay Program (Portugal) | 2007 - present | Teacher | ... | ... | ... |

Sources: Atkinson *et al.* (2009); Clotfelter and Ladd (1996); Eberts, Hollenbeck, and Stone (2002); Glewwe, Holla, and Kremer (2009); Glewwe, Ilias, and Kremer (2008); Ladd (1999); Lavy (2002, 2004); McEwan and Santibanez (2005); Mizala and Romaguera (2004); Muralidharan and Sundararaman (2008); Podgursky and Springer (2007); Rodriguez (2002); Santibanez *et al.* (2007); Schacter and Thum (2004); Springer, Ballou, and Peng (2008); Springer *et al.* (2008, 2009); Springer and Winters (2009); Taylor, Springer, and Ehler (2009); Vegas and Umansky (2005); Winters, Ritter, Greene, and Marsh (2009).

Knowledge-based pay and skill-based pay programmes are the most prominent type of incentive-pay policies that focus on rewarding educational inputs. Developed by a team of researchers with the Consortium for Policy Research in Education (CPRE) at the University of **Wisconsin**, these incentive plans induce teachers to obtain new skills related to classroom management and knowledge of curriculum development and curriculum content in core subjects. Several high-profile initiatives were launched in the **U.S.** in the late-1990s, including programmes in Cincinnati (**Ohio**), Douglas County (**Colorado**), Vaughn (**California**), and Washoe County (**Nevada**) (Milanowski, 2003).¹³

The acquisition of new knowledge and skills is believed to improve instructional effectiveness of a teacher and consequently student outcomes (Odden and Kelley, 1997). Input-based incentive systems are desirable because they are unlikely to cause “teaching to the test”, or to increase the chance of teachers focusing a disproportionate amount of effort toward those students already near some performance standard in hopes of

maximizing chances of earning a performance award. Firestone (1994, p. 550) further argued that these input-based pay programmes are “compatible with an outcomes focus because the resulting flexibility allows workers to do what is necessary to achieve those outcomes”.

Another appealing feature of input-based systems is rooted in the belief that the opportunity to pursue additional course work or to take on more responsibilities increases teacher interest and motivation (Odden and Kelley, 1997). This feature is very attractive if in fact it does find a way to balance the potential threat of external rewards reducing intrinsic motivation, particularly considering teachers are drawn to the profession by intrinsic factors (Kohn, 1999; Lortie, 1979; Rosenholtz, 1984). As noted by Johnson (2004, p. 46), “In itself, higher pay is unlikely to retain teachers – particularly the most able among them – if they find that they cannot attain the intrinsic rewards for which they initially entered teaching.”

Social psychological literature on public service motivation further argues that individuals seek to contribute to the public good in order to satisfy personal needs, not necessarily in response to incentives offered by organisations for performance or commitment (Courty, Heinrich, and Marschke *et al.*, 2005; Perry and Porter 1982; Rainey 1982; Wise 2004). Wise (2004), in particular, suggested that public-service motives had the potential to empower public servants to “overcome self-serving interests, moral inertia and risk avoidance” (as cited in Courty, Heinrich and Marschke, 2005), an effect to which Crewson (1997) attributed greater organisational commitment and lower employee turnover.

Despite proponents’ contention that the activities rewarded by knowledge- and skill-based pay programmes are associated with student achievement, a growing number of empirical studies indicate teacher effectiveness is, at best, marginally correlated with observable teacher characteristics such as teaching certificate held, level of education, licensing exam scores, and teaching experience beyond a few years (Goldhaber, Brewer and Anderson, 1999). For example, Goldhaber *et al.* (1999) found that observable teacher characteristics explain only about 3% of the differences in student achievement that are attributable to the teacher. In a related study, Goldhaber (2002, p. 50) argued that, “The evidence shows that good teachers make a clear difference in student achievement. The problem is that we really don’t know what makes a good teacher.”

Many stakeholders contend payment for output is likely to be a more efficient and productive way to remunerate teachers given variation in teacher effectiveness cannot be explained by activities rewarded in knowledge- and skill-based pay programmes. In the **U.S.**, for example, 97% of school districts remunerate teachers on criteria largely unrelated to schooling outcomes, while compensation payments for instructional personnel account for approximately 55% of K-12 current expenditures and 90% of instructional expenditures (Springer, 2009).

Input measures have been found to be related to student performance, notably observational evaluations that can be conducted by peers, supervisors, or external evaluators. Studies have reported that a standards-based evaluation can predict student achievement and can therefore provide performance evaluation scores as the basis for a performance-based incentive-pay programme or other personnel decisions (Kimball *et al.*, 2004; Milanowski, 2004). Additionally, Odden, Borman, and Fermanich (2004) reported observational evaluations not only provided a more “comprehensive model” of teacher effects on student achievement, but also have the potential to inform teacher practice by identifying specific areas of improvement.

However, the theoretical and empirical literatures on incentive-pay programmes in education identify a large number of concerns associated with subjective performance measures. Prendergast (1999) argued systematic errors, if frequently observed when employers rank workers, threaten the legitimacy of subjective assessment, often leading to leniency bias, centrality bias, and rent-seeking. Leniency bias is rooted in supervisors’ reluctance to rank the poorest performing workers, thereby distorting the overall distribution of rankings (Prendergast, 1999). Centrality bias, defined as the tendency for rankings to truncate the actual range of performance, introduces further

distortion into this distribution; supervisors may deliberately distort organisational standards and outcomes in order to redistribute rewards on the basis of personal preference (Adnett, 2003; Cutler and Waine, 2004; Marsden and Belfield, 2006). Finally, and as a function of these preferences, subjective assessment invites rent-seeking; workers focus effort away from critical tasks and toward influencing supervisors (Prendergast and Topel, 1993).

Further, raters' knowledge of prior performance appears to affect information processing by framing or anchoring current judgments (Huber, Neale and Northcraft, 1987). The order in which the observer sees good and poor performance therefore affects performance rating, with biased judgments about inconsistent extreme performance moving toward the prior general impression. Teams of raters may also attach greater weights to select elements than is appropriate (Schmitt, Noe and Gottschalk, 1986) or they may distort ratings for political reasons (Longenecker *et al.*, 1987).

The complex nature of teaching makes it difficult to identify specific criteria that should be present during an observational evaluation. While some have developed observational rubrics from supposed "research-identified" effective teaching behaviours (Danielson, 1996; Schacter and Thum, 2004), extensive training for raters is necessary. Observations are also very expensive as they require multiple observations on an individual teacher over the course of a school year.

In a recent study focused entirely on the predictive validity of supervisor evaluations, Jacob and Lefgren (2005) assessed the relationship between teacher performance ratings, as identified on a detailed principal evaluation, and teacher effects, as measured by student achievement gains. In estimating teacher effectiveness measures for 202 teachers of mathematics and reading in grades two through six, Jacob and Lefgren found a statistically significant and positive relationship between value-added measures of teacher productivity and principals' evaluations of teacher performance.

Another interesting dimension of this study was an "out of sample" prediction of 2003 student achievement scores based on principal ratings and teacher value-added estimates from 1998 through 2002. Students had higher average scores in mathematics and science if they had teachers with not only higher measured teacher effectiveness in prior years but also higher principal ratings. Jacob and Lefgren (2005) demonstrated further that the principal evaluation remained a statistically significant predictor of current student achievement even when teacher value-added (in the previous year) was included in the model. This finding suggests that principal evaluations provide an important independent source of information on teacher productivity.

Subjective evaluations also have the potential to serve as important formative assessments for teachers, although some have deemed subjective performance appraisals too fallible for high-stakes personnel decisions. If the teaching standards describe effective instructional strategies, as argued by Odden, Borman, and Fermanich (2004), then when these evaluative tools are implemented in the classroom, student achievement should increase. However, as summarised by Weisberg *et al.* (2009), teacher evaluation systems have failed to produce credible and accurate information about instructional practices of individual teachers.

Even though studies have indicated that principals are relatively adept at identifying above- and below-average teachers, it is unclear whether principal evaluations would persist in a high-stakes performance-pay programme. The fact that a principal identifies a teacher as "inadequate" on an anonymous survey does not mean necessarily that she will do so in a high-stakes environment. Take, for example, two interventions designed to increase teacher attendance rates in parts of **India** and **Kenya**. The first programme offered rather large bonuses if eligible teachers regularly came to work, as monitored by the school principal. An evaluation of the programme did not detect any discernible impacts on a number of teacher and student outcomes (Kremer and Chen, 2001). Analysis further revealed a majority of school principals distributed bonus awards to teachers, irrespective whether that teacher actually met the performance standards defined by the intervention.

In the second teacher attendance intervention, absence patterns were tracked using tamper-proof cameras with time and date functions, and treatment group teachers received a salary commensurate with the total number of days they attended work each month (Duflo, Hanna and Ryan, 2005). Treatment group teachers also received a modest bonus for each day they attended in excess of 21 days in a single month or an equally-sized pay deduction for each of the 21 days they did not show-up for work. Duflo *et al.* (2005) reported an average reduction in teacher absence rate of 19% and an increase in student performance on two standardised assessments.

In their comprehensive review of incentive-pay policies in developing countries, Glewwe, Holla and Kremer (2009) indicated the impact of the intervention evaluated by Duflo and colleagues may have been different from those reported in the Kremer and Chen (2001) study because the human element was removed from the monitoring process. The literature from outside the education sector further lends support in that managers often feel uneasy about monitoring colleagues in contexts with strong interdependencies, which suggests high-stakes performance evaluations conducted by personnel within the same school can possibly be unreliable and subject to perverse behaviours (Murphy and Margulies, 2004).

Educational outputs

Incentive-pay programmes have become increasingly focused on rewarding teachers or groups of teachers using outcome-oriented performance measures. Test scores measured on standardised assessments have gained popularity in large part because assessment instruments are routinely administered by most education systems, and standardised assessments represent an objective measure of performance. In an examination of professional compensation structures that are generally sympathetic to reform, for example, Bok (1993), as quoted in Ballou (2001), noted incentive-pay programmes could not work in the education sector because of the “inherently subjective and contentious process” of evaluating teachers.

Despite the considerable growth in output-focused measures of school, team, and teacher performance, these performance measures have received much criticism (Rothstein, 2009). Unlike sales or the billable hours of a doctor or lawyer, a teacher’s output is not measured readily in a reliable, valid and fair manner (Podgursky and Springer, 2007). Teaching involves a complex array of tasks that vary by the number of students taught, the grade levels of the students, the courses being taught, and the relationship of those courses to the tests (McCaffrey, Han and Lockwood, 2009). These complexities result in uncertainties about how teachers should be evaluated, and the way in which these uncertainties are resolved will ultimately affect the signals sent to teachers and other personnel (McCaffrey, Han and Lockwood, 2009).

Monitoring and rewarding individual teachers or groups of teachers according to student test scores captures only a fraction of the contribution of school personnel as well as the overall mission of a school. In the **U.S.**, for example, approximately two-thirds of teachers do not instruct in a tested grade or subject. Moreover, the typical student engages in a large number of academic and extra-curricular activities beyond subjects tested by most education systems. A sole reliance on objective measures, such as test scores, thus runs the risk of teachers focusing excessively on the metred, rewarded activity to the detriment of important dimensions of schooling (*i.e.* focusing on test-taking skills as opposed to creative or critical thinking skills). Additionally, cultivating and sustaining teacher support and buy-in can be problematic if a teacher believes the incentive-pay programme does not capture a significant aspect of their job or the specific performance objective is unclear (Murnane and Cohen, 1986). As a consequence, this lack of buy-in will weaken behavioural responses to the incentive programme (Chamberlin *et al.*, 2002).

Traditionally high- and low-performing students could be worse off in an incentive-pay programme excessively focused on output measures. Teachers may focus a disproportionate amount of their effort on the students who

are most likely to maximise the teacher's chances of earning an award. In the 1970s, for example, performance contracting experiments in the **U.S.** were found to be associated with teachers focusing excessively on students in the middle of the distribution to the detriment of their higher- and lower-performing peers (Gramlich and Koshel, 1975; Hannaway, 1996; Rapple, 1990). More recently, studies have examined a similar response to the 2002 No Child Left Behind (NCLB) Act (Ballou and Springer, 2009; Booher-Jennings, 2005; Neal and Schanzenbach, forthcoming; Reback, 2008; Springer, 2008).

Numerous other accounts from outside the **U.S.** have documented similar patterns of teacher behaviour in response to incentive-pay programmes. For example, in 1861, the **U.K.'s** New Castle Commission reformed the teacher compensation system as a means "to institute a searching examination... of every child in every school... and to make the prospects and position of the teacher dependent, to a considerable extent, on the results of the examination," (Bourne and MacArthur, 1970, p. 20, as cited in Chamberlin *et al.*, 2002). However, the incentive-pay programme ground to a halt some 30 years later, burdened by a legacy in which "teachers taught to the test, were confined to a narrow, boring curriculum, attempted to arrange the school intake, cheated, ignored bright children and drilled and beat the slower ones until they could satisfy the all-powerful inspectors," (Chamberlin *et al.*, 2002, p.32).

While incentive-pay programmes in the **U.K.** were linked with teachers focusing excessively on the metred, rewarded activity some 150 years ago, a much longer and more storied history of dysfunctional behavioural responses have also been documented. Implemented in 606 and lasting until about 1905, Imperial **China's** Keju examination system was largely used to identify the future employment and social status for each generation of students that completed formal schooling. However, according to Suen and Yu (2006), the Keju system suffered from construct-irrelevance and construct-underrepresentation which included rote memorization, an excessive focus on test-taking skills, and cheating in the form of nepotism, bribery, communicating with outside confederates during the examination, and students and their families hiring substitutes to complete the test. Although the Keju system is certainly an extreme example of potential negative responses to a high-stakes testing policy, the findings are relevant to the current discussion in that many of these negative consequences could have been reduced by broadening the performance measures.

A more recent example of manipulation comes from the International Child Support Incentive Program (ICSIP), a group incentive intervention that randomly assigned 100 schools in rural **Kenya** to either a treatment or control condition. A comprehensive evaluation completed by Glewwe, Ilias and Kremer (2008) revealed that the practices of teachers in the treatment group stayed relatively the same, apart from offering extra test preparation sessions outside of class time. The positive increase in the test scores of students enrolled in treatment group classrooms did not translate to non-tested subjects and were not sustained in subsequent years. An analysis of item-level test data further indicated that treatment condition students were significantly less likely to leave a test question blank.¹⁴

Additional problems are likely to surface when student and teacher interests are not well-aligned. While many standardised tests, such as upper-secondary school exit exams, are associated with high-stakes for students, there are many situations when exam performance has no bearing on the student (Bishop, 1996). In an incentive programme developed and implemented by **Israel's** Ministry of Education, teachers were measured by their students' performance on exams required for a certificate of matriculation, which has large financial implications for the students (Lavy, 2004). Even though teachers eligible for incentives led students to larger gains in achievement, the difference may not have generalised to situations where tests are high-stakes for students in alternating years: if a sixth grade student is required to pass in order to advance to the next grade but seventh grade scores do not carry the same weight, then the seventh grade teacher gains may exhibit bias due to decreased performance stemming from lower personal motivation.

Evaluating teacher performance solely on standardised test scores of students presents a non-trivial risk for teachers. In their seminal study of school accountability measures, Kane and Staiger (2001) noted:

The imprecision of test score measures arises from two sources. The first is sampling variation, which is a particularly striking problem in elementary [(i.e. primary)] schools. With the average elementary school containing only sixty-eight students per grade level, the amount of variation stemming from the idiosyncrasies of the particular sample of students being tested is often large relative to the total amount of variation observed between schools. The second arises from one-time factors that are not sensitive to the size of the sample; for example, a dog barking in the playground on the day of the test, a severe flu season, a disruptive student in a class, or favorable chemistry between a group of students and their teacher. Both small samples and other one-time factors can add considerable volatility to test score measures.

Volatility and other forms of error in the performance measure means teacher earnings are less certain because of factors beyond a teacher's control. Recognizing most teacher salaries are more predictable than in other professions (Liang, 1999), as the number of factors outside the control of teachers' increases, not only will the strength of the incentive be weakened but it could also lead teachers to demand higher levels of compensation. Thus, as summarised in a review of the economic complexities of incentive reforms, Asch (2005) noted that, "The optimum mix of the base and the risk component depends on the degree of risk aversion, earning variability, and the worker's cost of effort," (p. 316).¹⁵

Practitioners and policy makers interested in designing and implementing an incentive-pay programme need to know that even though VAM have gained a reputation for providing fair comparisons of teachers, this does not necessarily mean they can validly support such systems.¹⁶ McCaffrey, Han and Lockwood (2009) noted that VAM research to date has focused on the statistical properties of the measures from the perspective of methodological research rather than from the perspective of an algorithm that translates raw administrative data on students and teachers into information for managing human capital. Furthermore, to illustrate the many complexities and decisions that must be made when designing an incentive-pay programme, McCaffrey and colleagues also developed parameters for a series of hypothetical incentive-pay models and then test the performance of several Value Added modeling (VAM) strategies within this context.¹⁷

As VAM becomes increasingly popular in the education policy arena, it is critical that school systems understand the problems and limitations of their management information systems insofar as they relate to the creation of VAM. Common problems include mismatch of teacher and student data, multiple teachers with responsibility for the same student, unclear team or group composition, a lack of data and capacity to measure teacher value-added, and bonuses awarded to the wrong persons or denied to the right persons. These problems will be an embarrassment to the district and may jeopardise the credibility of the programme. Stakeholders should also bear in mind that even though the use of performance incentives is on the rise, the private sector has moved away from the sole reliance on quantitative measures of individual or group performance and begun to design incentive-pay programmes that incorporate both input- and output-based measures of employee performance (Rothstein, 2009).

Finally, poorly designed incentive-pay programmes can create perverse incentives whereby teachers move away from low-performing schools in order to maximise their chances of earning additional pay (Ladd, 1999; Neal, 2009), potentially exaggerating the already unequal distribution of teachers across historically advantaged and disadvantaged contexts. In a study of **North Carolina's** school accountability system, for example, Clotfelter, Ladd and Vigdor (2004) found the recruitment and retention of high-quality teachers even harder in low-performing schools.

Multiple measures – linking educational inputs and outputs

Schooling is often described as a multi-product, multi-purpose enterprise; the performance of a worker and the mission of a school comprise multiple dimensions. Both empirical and theoretical literature on incentive pay clearly indicate that if there is disconnect between an organisation's mission and the activity to which incentives are attached, employees may shift work toward the metred, rewarded activity, and away from other important activities (Dixit, 2002; Hannaway, 1992; Holmstrom and Milgrom, 1991). Many advocates argue that the use of multiple measures in an incentive-pay programme means the performance of schools, teams of teachers, or individuals can be monitored more effectively at various points throughout the school year, all of which can reduce the chances for system-gaming.

A large number of studies have documented how narrowly-focused performance measures can create greater opportunity in the long run for employee cheating and opportunistic behaviour. In an incentive-pay programme that rewards teacher performance solely on student test scores, dysfunctional behaviours can take the form of changing student responses on answer sheets, providing students with correct answers during the examination, or obtaining copies of test material prior to administration in order to prepare students (Jacob and Levitt, 2003).¹⁸

Although linking educational inputs and outputs can reduce the likelihood of system-gaming, there can be significant costs associated with an incentive-pay programme that evaluates teachers across multiple performance measures. Guthrie and Prince (2009) reported that both Philadelphia (**Pennsylvania**) and Steamboat Springs (**Colorado**) cancelled incentive-pay programmes after learning of the cost of implementing the teacher performance appraisal system. Similarly, several school districts in **Florida** were interested in participating in the state pay-for-performance programme, but abandoned the programme after learning the cost of administering the battery of student assessments required to participate (Guthrie and Prince, 2009).

Each of the stakeholders involved in the design and implementation of an incentive-pay programme may advocate for integrating unique performance measures that satisfy their personal interest or that of their constituents. For example, parents expect schools to “foster creativity, curiosity, self-esteem, tolerance, good citizenship, athletic performance, and a host of other objectives” (Hannaway, 1996, p. 103), which may correspond imperfectly to the goal of schooling as defined by a policy maker. Even though adopting multiple measures can go a long way toward capturing the complexity of the schooling process, stakeholders need to be cognisant of the potential for conflicting goals engendering the misallocation of resources (Baker, 1992, 2002; Banker and Datar, 1989; Feltham and Xie, 1994; Holmstrom and Milgrom, 1991).

Similar to a hybrid unit of accountability, performance evaluations using multiple measures require that weights be assigned to each dimension of the system, an aspect of incentive-pay programmes that has tended to be ignored in the education sector-specific literature. Programmes like **New York City's** School-Wide Performance Bonus Program placed the greatest weight on student test scores (30% student achievement levels and 55% student achievement gains), while **Mexico's** *Carrera Magisterial* adopted a system whereby 70% of a teacher's evaluation relied on input-based measures such as seniority, academic degree, and professional preparation. Stakeholders will certainly request an explanation as to why various criteria are weighted the way they are, and these weights, for better or worse, can send a strong signal to students, parents, educators, and policy makers about the value programme designers placed on specific aspects of the schooling process.

While performance measurement systems that incorporate a myriad of metrics are likely to reduce unintended consequences by offering a more balanced evaluation system, at present, a considerable number of school systems do not have the organisational capacity to implement a programme incorporating multiple measures of teacher, team, or school performance. Even the most basic of incentive-pay programmes requires that a significant amount of personnel time be allocated to monitoring implementation, responding to and resolving

inconsistencies in the design of the programme, and addressing questions and complaints submitted by school personnel and other stakeholders.

In total, contemporary incentive-pay programmes have a lot of room for improvement when it comes to identifying the optimal mix of input- and output-focused performance measures and assigning weights to each of the evaluation criteria adopted. We recommend that school systems need to rely on rigorous, formative evaluations of their incentive programmes as trial and error will need to inform programme development and design in both the short- and long-run (Courty and Marschke, 2003; Podgursky and Springer, 2007). Moreover, rushing implementation before teachers and other school personnel understand how the programme works will in all likelihood set the stage for failure.

PERFORMANCE STANDARDS AND THRESHOLDS

Performance standards and thresholds determine the required level of performance for a school, team of teachers or individual teacher to secure a reward. The standards and thresholds defined by the incentive-pay programme dictate the number of units that can earn a bonus as well as what scale or minimum standards these units must meet. Our discussion reviews three forms of incentive-pay models, including linear models, step functions, and limited linear models. We also discuss the fine balance that must be established between performance thresholds and the number of units projected to earn a bonus award.

Linear models

Linear models, also known as continuous standards, refer to rewards based on the average level of or average gain in, desired outcome (Lazear, 2003). In school settings, an example using change in test scores is outlined by Lazear (2003), where β represents the relationship between earnings and test scores, and K is a constant that sets salaries at their appropriate level.

$$\text{Teacher's salary} = N(\Delta \text{Score})\beta - K$$

Linear models have several attractive features. Linear models are sufficiently robust, do not depend on the timing of the agent's information, and do not require all available information to make a bonus determination (Holmstrom and Milgrom, 1987). Another important feature is that linear models value progress at all parts of the achievement distribution (Lazear, 2003). Instead of focusing on students who are close to a certain cutoff point, a linear model allows for the aggregate growth of all students to determine bonus eligibility.

Countering arguments in favour of linear models, Kole (1997) demonstrated that an incentive-pay programme using a standard broad categorization ignores many important aspects of compensation contracts and can understate incentives. Her findings suggest that exogenous factors that influence performance may bias results if they, for example, impact high- or low-achieving students differentially. Furthermore, linear systems may present more of a logistical challenge for systems due to greater complexity of bonus calculations. Increased differentiation of salary among employees may require bonus determination at an individual level, a requirement that could place undue burden on human resources personnel.

Step functions

Step functions, or non-continuous models, distribute rewards based on meeting a minimum standard or threshold. While multiple standards or steps are possible, step functions establish a cutoff point below which either no bonus is awarded or penalties are incurred. Step functions are optimal when the output is very sensitive to the effort put forward by the unit of accountability in the general vicinity of the performance threshold (Holmstrom, 1979; Mirrlees, 1999). In other words, step functions create strong incentives for schools, teams of teachers, or individual teachers facing a significant reduction in pay if they reduce their effort even marginally.

The literature on incentive contracts has demonstrated that step functions are too easily manipulated, especially when they cover a fixed period of time (Dixit, 2002). In such a context, step functions provide comparatively weak incentives for effort either after the threshold is reached or after it becomes unattainable (Asch, 1990; Holmstrom and Milgrom, 1987; Oyer, 1998). For example, in a recent study of physician medical groups, Mullen, Frank and Rosenthal (2009) found that physicians' groups with baseline performance at or above the performance threshold for receipt of a bonus improved the least, but garnered the largest share of the bonus payments.

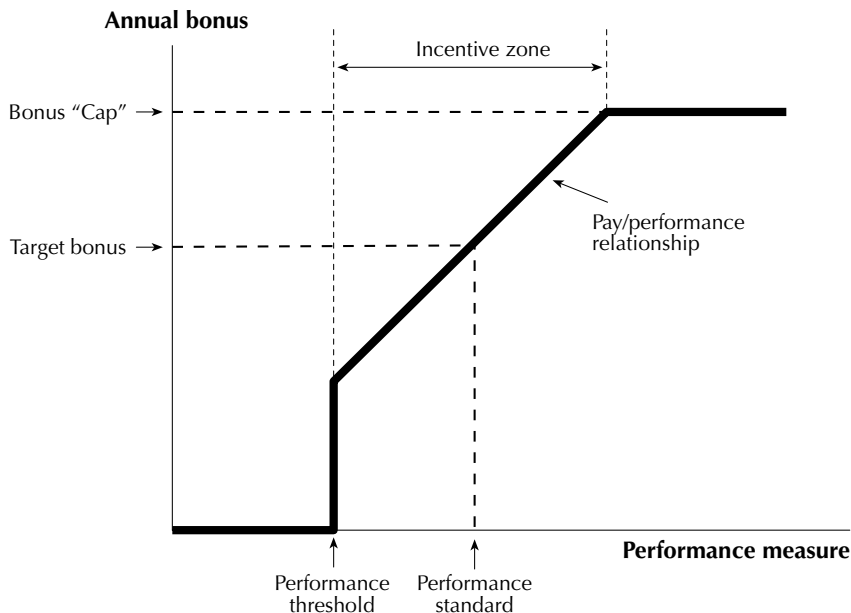
Limited linear models

It is possible to combine elements of both linear and step function models in an incentive programme using a limited linear model framework. As outlined by Murphy (2001), Figure 3.2 displays the theoretical relationship between a performance measure (x-axis) and the annual bonus (y-axis) within the context of a limited linear model. The baseline performance threshold establishes the minimum possible level of performance associated with a bonus award. Beyond this performance threshold floor, there is a positive linear relationship between increasing performance and an annual bonus award.

A limited linear model is further characterised in Figure 3.2 by the fact that at some point along performance continuum the maximum size of a bonus award is capped. Any increase in the performance of the unit of accountability beyond this predetermined level does not earn a school, team of teachers, or individual teacher any more money. The range of performance rewarded under an incentive-pay programme is called the incentive zone, and midway between the lower- and upper-performance thresholds is a predicted level of performance and corollary bonus that can be used in budget projections.

Figure 3.2

Performance standards and thresholds



Note: Adapted from Murphy (2001).

Several aspects of a limited linear model are particularly attractive. The model allows for improvement in the performance of the unit of accountability to increase the size of the annual bonus awarded at any point in the incentive zone. The limited linear model also restricts the bonus size of the lowest performers, thereby allowing for the possibility of the incentive-pay programme to send a strong signal to those at the bottom of the performance continuum. Finally, the financial risk typically imposed by extremely high performance is reduced by the bonus cap.

Threshold levels

Another consideration for thresholds is the targeted number of schools, teams of teachers, or individuals that designers of an incentive-pay programme expect to obtain a bonus award. Inherent in this decision is the balance between attainability and exclusivity, which not only plays a vital role in eliciting behavioural responses, but also dictates whether an incentive-pay programme will realise its full potential.

Reward programmes in **Bolivia** and **England** provide examples at each extreme. On the one hand, in Bolivia's merit wage programme, only 2% of teachers performed sufficiently on content knowledge exams to obtain a reward during the first year of the programme. While public outcry subsequently led to slightly more than 25% of teachers earning a reward in the second year of the merit wage programme, continual resistance on the part of teachers led to dissolution of the programme (Mizala and Romaguera, 2004).

In **England's** Performance Threshold System, on the other hand, about 88% of those teachers eligible to participate in the incentive-pay programme elected to do so. Among participating teachers, an astounding 97% met the predetermined standard established for earning a bonus award. Ultimately, the low level of perceived value and worth of attaining a performance award was attributed with the weak motivational influence elicited by the System (Atkinson *et al.*, 2009).

SIZE AND DISTRIBUTION OF BONUS AWARDS

The size of bonus, or payout level, refers to the amount of the total bonus award a school, team of teachers, or individual can earn. Distribution relates to the share of teachers that received a bonus award and how bonuses vary among employees. While no clear guidance exists on the optimal size of a bonus in a teacher incentive-pay programme, several studies suggested the size of bonus awards for teachers have been so small that the motivational value of most incentive systems have been compromised (Chamberlin, *et al.*, 2002; Heinrich, 2007; Malen 1999; Taylor and Springer, 2009).

In this section, we first discuss various topics found in the theoretical and empirical literature in an effort to offer a broad perspective on the size of bonuses and distribution of bonus awards. We then provide some general information on non-monetary incentives, which can take the form of promotion and advancement, positive feedback, public or private recognition, decreased regulations or requirements, increased status, greater decision-making capacity, access to in-house education or formal training, and so on. While the mechanism may differ from monetary rewards, the underlying principle of influencing behaviour remains constant.

Size of bonus

Table 3.2 displays summary statistics on the size of bonus awards for a number of incentive-pay programmes.¹⁹ In the **U.S.**, the size of the bonus award at the programme level ranges from a low of 0.4% to a high of 365% of a teacher's monthly salary. Both of these estimates come from the **Texas** Educator Excellence Grant programme. Estimates from international incentive-pay programmes are displayed in the bottom half of Table 3.2, where the smallest bonus award was associated with **Chile's** *Sistema Nacional de Evaluación de Desempeño de los Establecimientos Educativos* (SNED) programme. **Mexico's** *Carrera Magisterial* programme generally awarded the largest bonuses, which in some cases exceeded 200% of a teacher's monthly salary.

Table 3.2

Performance measures in incentive pay programmes

| | Period | Performance Measures | |
|--|----------------|---|---|
| | | Educational Inputs and Processes | Educational Outputs |
| United States | | | |
| School Incentive Program (Dallas, Texas) | 1992 - 1995 | Student attendance | Student achievement levels; Student test score gains; Student retention rates |
| Merit Pay Program (State of Michigan) | 1996 - 1997 | Student evaluation of teacher | Student retention rates |
| Achievement Challenge Pilot Project (Little Rock, Arkansas) | 2005 - 2007 | ... | Student test score gains |
| Teacher Advancement Program (United States) | 1999 - present | Professional development; Career ladder (three levels); Supervisor evaluation | Student test score gains (group- and school-level) |
| Governor Educator Excellence Grant Program (State of Texas) | 2005 - 2008 | Student attendance; Teacher attendance | Student achievement levels; Student test score gains |
| Texas Educator Excellence Grant Program (State of Texas) | 2007 - present | Student attendance; Teacher attendance | Student achievement levels; Student test score gains |
| Professional Compensation System for Teachers (Denver, Colorado) | 2005 - present | Knowledge and skills; Professional evaluation; High-need schools; Hard-to-staff subjects | Student test score gains; School performance |
| ASPIRE Program (Houston, Texas) | 2005 - present | Teacher attendance | Student test score gains (group- and school-level) |
| Team Incentive Experiment (Round Rock, Texas) | 2008 - present | ... | Student test score gains |
| Project on Incentives in Teaching Experiment (Nashville, Tennessee) | 2006 - 2009 | ... | Student test score gains |
| Q-Comp (State of Minnesota) | 2006 - present | Career advancement; Professional development; Professional evaluation; High-need schools; Hard-to-staff subjects | Student achievement levels; Student test score gains |
| School-Wide Performance Bonus Program (New York City, New York) | 2008 - present | Student attendance; Student, parent and teacher perception of school learning environment | Student achievement levels; Student test score gains |
| International | | | |
| School Performance Program (Israel) | 1996 - 1997 | ... | Number of credit units earned per student; Student receiving matriculation certification; School dropout rate |
| International Christelijk Steunfonds Incentive Program (Buso and Teso, Kenya) | 1998 - 1999 | ... | Student test score gains; Student achievement levels |
| Teacher-Incentive Experiment (Israel) | 2001 | ... | Student achievement levels |
| Randomized Evaluation Project (Andhra Pradesh, India) | 2006 - present | ... | Student test score gains |
| Carrera Magisterial (Mexico) | 1993 - present | Seniority; Academic degree; Professional level; Professional preparation; Professional performance | Student achievement levels |
| Salario al Mérito (Bolivia) | 1998 - 1999 | Teacher content knowledge | ... |
| Incentivo Colectivo a Escuelas (Bolivia) | 2001 - present | School organization; Teacher training and retention; Regularity in school management; Number of students per class; Student-teacher ratio; Teaching initiatives developed by teachers; Parent participation | Student pass rates; Student drop-out rates; Student repetition rates |
| Performance Threshold System (England) | 1999 - present | Supervisor evaluation | Student test score gains |
| Sistema Nacional de Evaluación de Desempeño de los Establecimientos Educativos (Chile) | 1996 - present | School initiative; Improved working conditions | Student test score gains; Student achievement levels; Student pass rates; Student retention rates |
| Plan de Estímulos a la Labor Educativa Institucional (El Salvador) | 2000 - present | School management; Educational management; Institutional planning; Teacher management. | Student drop-out; Student retention rates |
| Performance Pay Program (Portugal) | 2007 - present | Teacher attendance; Attendance at training sessions; Management and pedagogical duties; Involvement in research projects; Parent feedback about teacher | Student test scores |

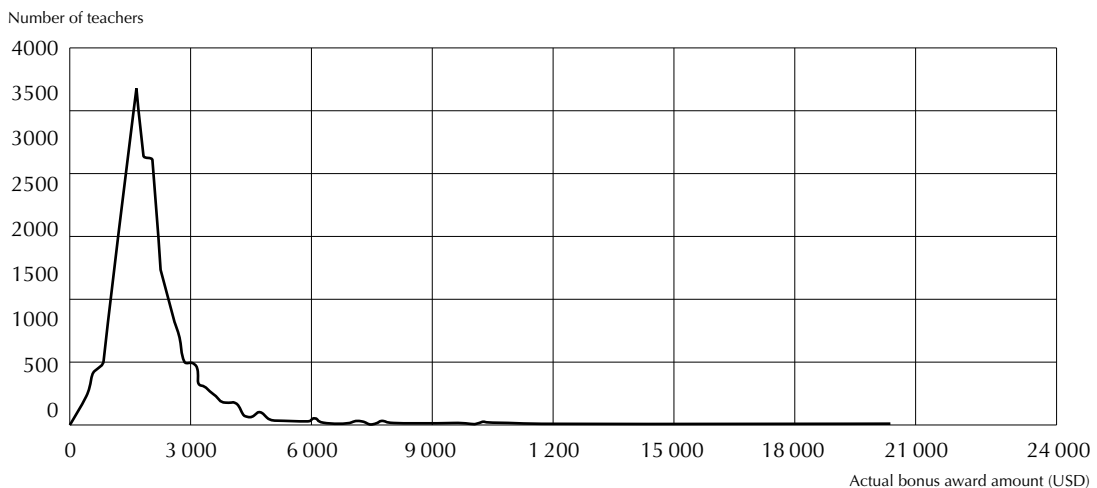
Sources: Atkinson et al. (2009); Clotfelter and Ladd (1996); Eberts, Hollenbeck, and Stone (2002); Glewwe, Holla, and Kremer (2009); Glewwe, Ilias, and Kremer (2008); Ladd (1999); Lavy (2002, 2004); McEwan and Santibanez (2005); Mizala and Romaguera (2004); Muralidharan and Sundararaman (2008); Podgursky and Springer (2007); Rodriguez (2002); Santibanez et al. (2007); Schacter and Thum (2004); Springer, Ballou, and Peng (2008); Springer et al. (2008, 2009); Springer and Winters (2009); Taylor, Springer, and Ehler (2009); Vegas and Umansky (2005); Winters, Ritter, Greene, and Marsh (2009).

Table 3.2 presents information that suggests the magnitude of the maximum bonus award has increased noticeably over time. Lower-secondary school mathematics teachers participating in a pay-for-performance experiment in Nashville, **Tennessee**, conducted by the National Center on Performance Incentives, set bonus awards as large as an additional USD 15 000 per year, or the equivalent of 280% of the average monthly salary. The National Center on Performance Incentives is also evaluating a teacher incentive experiment that awards the top one-third of treatment teams with a USD 6 000 bonus per teacher. Casual empiricism suggests the magnitude of the maximum bonus increased over time not only in response to past reform efforts that provided unappealing award amounts, but also to increased interest in learning whether teachers responded to substantial bonus awards, even if the incentive system lacked a complete array of measures.

Figure 3.3 displays how often teachers participating in the Governor Educator Excellence Grant (GEEG) programme received a bonus award, by bonus award amount (Springer *et al.*, 2008).²⁰ Bonus awards ranged from a low of USD 20 to a high of USD 20 462, with most teachers receiving between USD 1 000 and USD 3 000. What is most striking about the information is that nearly 90% of the teachers that received a bonus award earned less than USD 3 000, despite programme guidelines that encouraged teachers and schools to offer bonuses between USD 3 000 and USD 10 000 (Springer *et al.*, 2008). The pattern in Figure 3.3 suggests most schools approved incentive-pay programmes with a relatively large number of small bonus awards, which, according to Taylor and Springer (2009), did not appear to have induced any significant changes in teacher productivity.

Figure 3.3

Distribution of actual awards in the Governor Educator Excellence Grant programme



Source: Adapted from Springer *et al.* (2008).

Distribution of award

Bonus award distribution systems determine how evenly an incentive-pay programme distributes rewards to eligible employees. An egalitarian distribution plan distributes incentive money widely, in contrast to plans that reward larger sums of money to fewer schools, teams of teachers, or individuals. There is no clear guidance, however, whether an incentive-pay programme should reward a large number of relatively small awards to

teachers in a school, or reward a smaller number of teachers with a relatively large award (Taylor, Springer and Ehlert, 2009). Our discussion regarding the distribution of bonus awards therefore draws heavily on the general personnel incentive literature.

Proponents argue that individualist reward plans help create a meritocracy able to retain an organisation's highest performers, attract similar talent over the long run, send a clear signal to the lowest performers to improve or move elsewhere, and are more cost-effective (Ehrenberg and Smith, 1994; Milgrom and Roberts, 1992; Pfeffer and Langston, 1993; Zenger, 1992). At the same time, a growing body of research suggests egalitarian distributions promote cooperation and group performance, which are critical in participative organisations. Milgrom and Roberts (1992) suggested, moreover, that greater pay dispersion may elevate the performance of the lowest performers.

In a thorough review of the characteristics of incentive-pay programmes that have been adopted in **Bolivia, Chile, El Salvador, and Mexico**, Mizala and Romaguera (2004) reported that Bolivia's *Incentivo Colectivo a Escuelas* (ICE) awarded approximately 5% of eligible teachers a bonus ranging between 5% and 19% of annual income. Chile's SNED distributed a modestly larger number of bonus awards but the amount of those awards was slightly smaller in size, while Mexico's CM awarded the largest sums, with bonuses ranging between 27% and 224% of annual income (Mizala and Romaguera, 2004). In comparison to the ICE and SNED programmes offering bonuses to about 5% and 27% of eligible teachers, respectively, the CM incentive programme provided bonus awards to approximately 75% of eligible primary school teachers (Mizala and Romaguera, 2004).

A study of the Governor's Educator Excellence Grant programme in **Texas** examined the proposed distribution of awards to school personnel within schools participating in the programme. Taylor, Springer and Ehlert (2009) found that, on average, the proposed distribution of awards was more equal than the distribution of disposable income in the **U.S.**, but less equal than the distribution of teacher salaries in the schools. Acknowledging the fact that virtually all schools participating in the GEEG programme adopted highly egalitarian award distribution schemes, as well as the programme not having any discernible association with teacher productivity, a subsequent study of the programme suggested that when given the opportunity, teachers appear to design relatively weak incentive systems (Taylor and Springer, 2009).

Stakeholders interested in designing an incentive-pay programme need to consider the type of educational institution when establishing the size and distribution of bonus awards. For example, in a study of the relationship between wage dispersion and productivity in institutions of higher education, Pfeffer and Langston (1993) found less egalitarian salary structures decreased productivity and increased dissatisfaction among faculty. Yet, the negative effects were not as apparent in private colleges and universities, which the authors attributed to the fact that employee pay in most private institutions is not necessarily a matter of public record.

Newspapers, blogs, and other media outlets can obtain information on the distribution of bonus awards through the Freedom of Information Act (FOIA) in the U.S. Several newspaper outlets have obtained teacher level bonus information and then created venues for the general public to access and search this information via the Internet.²¹ The way in which information about wages is communicated to employees can greatly affect their attitude toward the programme and behaviour (Gerhart and Milkovich, 1990). We recommend stakeholders not only communicate regularly with the media, but also engage and work with teachers and other school personnel so they are well informed of the potential for publication of their bonus award amount.²²

Non-monetary incentives

While this chapter has focused primarily on programmes involving monetary incentives, non-monetary incentives can also function as workplace motivators. Many schools and school systems nominate and recognise certain teachers who perform above expectations, or offer teachers improved working conditions, unpaid leave,

mentoring and induction programmes, and job enlargement. In fact, in a review of reform strategies adopted in response to the problem of rural teacher recruitment among more than 20 developing countries, McEwan (1999) identified special training programmes, reduction in seniority requirements before promotion, and longer vacations as the most prevalent forms of non-monetary incentives.

Stigma and penalties have also been identified as potentially high-powered non-monetary incentives. Several states in **Germany** developed a reporting programme that publicised student learning gains association with schools, which Juerges, Richter and Schneider (2004) reported was a positive influence on the quality of teaching and student outcomes because teachers wanted to preserve and protect their reputation. Stigma and threats of sanction have also been found to have mostly positive effects in **Florida** (Chiang, forthcoming; Figlio and Rouse, 2006; West and Peterson, 2006). At the same time, while some stakeholders argued that information provided by the German reporting system should be used to inform personnel decisions (*i.e.* hiring, firing, and promotion/tenure), two national strikes organised by the teacher union resulted in political leaders in Germany committing to not use the results in this way (Juerges, Richter and Schneider, 2004).

PAYOUT FREQUENCY

Payout frequency refers to the rate of award distribution as well as the time interval between assessment of the incentivised activity and distribution of the performance award. The literature in psychology suggests incentives are most effective when the unit under observation is awarded consistently with minimum time between action and reward (Skinner, 1981). While some delay in gratification is possible, immediacy minimises the difficulty in association between an individual's behaviour and reward. Further, expectancy theory notes more frequent distribution and minimal temporal delay solidify connections between outcome and reward (Vroom, 1964). The delay that now occurs between teacher performance and payment of premium is longer than any behavioural psychologist would suggest for purposes of reinforcing good actions.

Most incentive-pay programmes in the education sector distribute awards on an annual basis, corresponding to the academic year, often due to the timing of results from end-of-year assessments. While practical, minimizing the time interval by tracking progress throughout the year may promote a more transparent link between action and reward. Further, assessing and rewarding behaviours more frequently during the school year will allow for consistent reinforcement and likely result in a more pronounced influence on behaviour (Hollensbe and Guthrie, 2000). This connection is present in the business sector; research conducted by the Consortium for Alternative Reward Strategies found compensation plans with greater payout frequency were linked to business performance improvements (McAdams and Hawk, 1992).

Eberts, Hollenbeck and Stone (2002) studied the effect of an incentive scheme in an alternative upper-secondary school in **Michigan** that rewarded teachers at multiple points throughout the school year. The programme was designed to address a growing dropout rate problem, and introduced a bonus system that paid teachers to raise course completion rates of students at risk of dropping out. Although the programme relied on a rather limited set of performance indicators, the intervention teachers were paid a 12% bonus each quarter if at least 80% of students from the previous quarter were retained.

Increasing the payout frequency may also decrease the likelihood of system gaming since behaviours are assessed on a more regular basis. As found in a study of Navy recruiters, for example, when their performance was evaluated at 12-month intervals, the recruiters' productivity significantly increased as the evaluation date approached. This spike in performance was characterised by a sizeable reduction in output shortly after the performance evaluation. Similarly, Courty and Marschke (1997) reported that training centers under the Job Training Partnership Act focused efforts on more qualified participants entering the programme, and the timing of their graduation, to meet certain standards.

CULTURAL CONSIDERATIONS

The impact of an incentive-pay system may reflect characteristics inherent in cultural contexts. Cultural dimensions, as described in Hofstede (1980, 2001), provide categories in which countries can be compared regarding reward preferences, including power distance, collectivism versus individualism, femininity versus masculinity, and uncertainty avoidance. Power distance refers to how people feel power should be distributed as well as how those with power should be viewed: a high power-distance environment would say power should be limited to fewer individuals whose decisions should be accepted, while lower power distance would have more people in control and would allow the challenging of leaders. A collective culture would place the allegiance to the group above that of the individual with this being reversed in an individualist culture. A feminine culture would value caring for others and establishing social support while a more masculine society would place a higher value on achievement, the accumulation of wealth, and assertiveness. Uncertainty avoidance refers to the extent a culture avoids situations that are ambiguous or have unpredictable outcomes.

Reward preferences can be directly linked to the chance of an incentive-pay programme actually meeting individual needs and cultural values found within particular countries. In their case study of four countries that varied on the dimensions established by Hofstede (1980, 2001), Chiang and Birtch (2005) found that countries with high masculinity scores were more likely to value material rewards as compared to those countries with more feminine scores. They found a similar preference for financial rewards between individualistic and collective cultures, while responses to incentive-pay programmes significantly diverged according to cultural factors. Even though the evidence reported by Chiang and Birtch (2005) is based on observations from a limited number of countries, their findings support the argument that cultural influences are an important aspect for education practitioners and policy makers to consider when designing and implementing an incentive-pay programme.²³

Furthermore, in Table 3.3 as defined by Hofstede (1980, 2001), we illustrate the power dimension scores in several countries with incentive programmes to further assess the unique nature of the cultural context of each incentive-pay programme. Take, for example, the programme designed and implemented by **Israel's** Ministry of Education, which was considered highly competitive because the incentive structure relied on a rank order tournament structure; that is, there were a limited number of people who could earn the reward and teachers eligible for a bonus award were knowingly in direct competition with one another. A plausible explanation for the acceptance of the programme by school personnel, while similar incentive programme designs have typically failed in other contexts, is that Israeli culture has a comparatively high individual score. While more research is definitely needed to determine specific relationships between design features of incentive-pay programmes and the cultural dimensions displayed in Table 3.3, it is important to consider the cultural context when designing an incentive-pay programme.

Table 3.3

Cultural dimensions for selected countries with incentive pay programmes

| | Power Distance | Individualism | Masculinity | Uncertainty Avoidance |
|-------------------------|----------------|---------------|-------------|-----------------------|
| India | 77.0 | 40 | 48 | 56 |
| Israel | 13.0 | 81 | 54 | 47 |
| England (Great Britain) | 35.0 | 35 | 89 | 66 |
| Kenya (Eastern Africa) | 64.0 | 52 | 27 | 41 |
| Chile | 63.0 | 23 | 28 | 86 |
| Mexico | 81.0 | 30 | 69 | 82 |
| World Average | 56.5 | 43 | 51 | 65 |

Note: Adapted from Hofstede (1980, 2001).

CONCLUSION

In this chapter we reviewed and synthesised many of the theoretical and empirical arguments connected with various design elements found in incentive-pay programmes. We focused on specific design elements including the incentive structure, unit of accountability, performance measures, performance targets and thresholds, size and distribution of bonus awards, and payout frequency. We also discussed the potential relationship between cultural dimensions and the design of incentive-pay programmes.

Incentive structures can impact the number of teachers likely to receive a bonus award as well as the required financial commitment needed to implement and sustain the incentive-pay programme over time. Rank ordered tournaments clearly limit the financial exposure of a fixed performance contract, which is advantageous considering the volatility and noise inherent in many performance measures found in the education sector. School personnel and organisations have historically resisted tournament incentive structures, however, because they can foster competition and other conditions that are incompatible with the norms of the teaching profession.

Incentive-pay programmes have overwhelmingly conceptualised the unit of accountability as being either at the individual- or school-level, or a combination of the two. Yet, this conceptualization misses a potentially important unit found within schools: grade-level or departmental teams of teachers or any other within-school unit. Education stakeholders should consider more closely within-school teams as the unit of accountability, which several theoretical and empirical studies from outside the education sector indicated have the potential to combine the cohesive benefits of group accountability with the individual belief that productivity directly impacts bonus eligibility. The education system should also continue to explore more complex hybrid models that may be more suitable in multi-product, multi-purpose environments.

Careful consideration needs to be given to how the performance of schools, teams of teachers, or individuals is measured. A large literature indicates employee behaviour will tend to shift in the direction of incentivised actions, and both potential positive and negative outcomes should be anticipated. Multiple measures can often minimise gaming behaviour and provide a more valid, reliable, fair and comprehensive assessment of teacher, team and school effectiveness, all of which are crucial to the motivational effect of the incentive-pay programme. Furthermore, there are reports of more and more private sector firms adopting both input- and output-based measures when assessing employee performance.

In regards to performance standards and thresholds, incentive-pay programmes have usually adopted either a linear or step function model. Most researchers have advocated for linear models due to design simplicity and their robust nature, even though some of the behavioural economic literature cautions that a linear model ignores important information about the actual production process. Given the relatively sparse literature on performance standards and thresholds, it is important for future evaluation research and development work to closely examine these structures in an effort to further inform stakeholders of the relative merits and weaknesses of each system.

The size and distribution of incentive-pay awards have tended to vary dramatically among programmes, although a few relatively recent studies have suggested that teachers may be predisposed to adopting highly egalitarian award distribution schemes. When designing an incentive-pay programme, it is critically important to keep in mind that bonus awards need to be sufficiently large to elicit behavioural responses and not so large that gaming or cheating becomes a concern. We further recommend that incentive-pay programmes in the education sector increase the frequency in which performance is evaluated and awards are distributed. To the point that they are feasible, multiple evaluations and performance objectives or benchmarks during the school year strengthen the connection between behaviour and reward. Such arrangements may help teachers, teams of teachers, or schools to monitor and alter behaviour at several points in time instead of basing their performance bonus on a one-time annual assessment.

Cultural norms and preferences toward design elements of incentive-pay programmes are likely important considerations during the design and implementation phase. While a rank-order tournament evaluating individual teacher performance may have elicited support and had a positive impact on student achievement and teacher behaviour in **Israel**, countries predisposed toward more collective cultural processes and norms might be more successful adopting a group-level or hybrid model using a fixed performance contract incentive system. Even within a country, regions may also differ in socio-cultural characteristics as well as the economic situations that influence the ideal design of incentive systems.

It is clear the literature on the design components of incentive-pay programmes in the education sector still needs to be developed before we can define how the first or second-best system should be designed. Educational research, practice, and policy needs to continue to support policy innovation combined with rigorous, independent evaluations of both the short- and long-run impacts on student achievement, teacher attitudes and behaviour, and organisational dynamics.

NOTES

1. Incentive-pay programmes have grown in popularity. High-profile programmes have included **Bolivia's** Merit Wages and ICE program (Vegas and Umansky, 2005; Mizala and Romaguera, 2004); **Chile's** *Sistema Nacional de Evaluación del Desempeño de los Establecimientos Educacionales* (SNED) (Mizala and Romaguera, 2003); **El Salvador's** school awards program (Rodríguez, 2002); **Mexico's** *Carrera Magisterial* (McEwan and Santibanez, 2005; Santibanez *et al.*, 2007); **New York City's** School-Wide Performance Bonus Program (Springer and Winters, 2009); **Texas'** Governor's Educator Excellence Award Programs (Springer *et al.*, 2008; Springer *et al.*, 2009); the **U.S.** Teacher Incentive Fund (Podgursky and Springer, 2007); programmes developed by **Israel's** Ministry of Education (Lavy, 2002; 2007); and experiments in Andhra Pradesh, **India** (Muralidharan and Sundararaman, 2008) and in the Busia and Teso districts of western **Kenya** (Glewwe, Ilias, and Kremer, 2008). In the **U.S.**, pay-for-performance experiments have recently been implemented and are currently being evaluated in Nashville, **TN**, New York City, **NY**, and Round Rock, **TX**.

2. In response to the adoption of **Portugal's** performance-pay policy, there were two national strikes led by teachers and their unions (Martins, 2009).

3. While at first sight there appears to be a significant increase in student learning, the percentage of schools meeting the proficiency standard may not in fact represent a true increase in student achievement. There is a healthy literature on weaknesses of high-stakes assessments and test-based accountability systems, including test scores becoming artificially inflated because a state may have adopted a new standardised assessment, recalibrated proficiency standards, or repeatedly administered a similar test form from one school year to the next (see, for example, Linn, 2000; Koretz, 2003; 2008). In describing what is widely-known as the "Lake Wobegon Effect" [This may need an explanation for the international audience](e.g., scenario where schools or teachers receive overly positive ratings), Cannell (1987; as cited in Linn, 2000) argued that "standardized, nationally normed achievement tests give children, parents, school systems, legislatures, and the press inflated and misleading reports on achievement levels" (p. 7).

4. In the **U.S.**, it has been argued that the increased compression of the federal sector's compensation system relative to that in the private sector has made it more difficult for the federal sector to recruit and retain high-quality employees (Gibbs, 2001; Borjas, 2002; Asch, 2005).

5. Work being conducted by researchers at the National Center on Performance Incentives has begun to explore many of the operational complexities associated with incentive-pay programmes and the effect these complexities have on measures of teacher and team performance.
6. Control condition refers to those classrooms not exposed to one of the four interventions.
7. An alternate draw is a selection process where the leader of the first team selects from a pool of eligible candidates and then the leader of the next team selects from the remaining pool and so on until all eligible candidates are selected.
8. Independent evaluations have been conducted by Springer, Ballou, and Peng (2008) and are currently being conducted by Glazerman, Allison, McKie, and Carey (2009).
9. Levin and McEwan (2001) offer a very thorough and practical overview of cost-effectiveness analysis when evaluating or considering educational programmes and policies.
10. The Center for Educator Compensation Reform offers a number of useful resources regarding the implementation of incentive-pay programmes. For more information, visit <http://www.cecr.ed.gov>
11. Previous research typically codes subjective evaluations of teacher performance as an input measure, output measure, or process measure. The same holds true for student and teacher attendance. In this review, we consider these process measures.
12. Even though the vast majority of states in the **U.S.** have funded incentive plans around teacher recruitment and retention, as observed by Loeb and Miller (2007), most of these programs are not well-aligned with teacher labour market realities, nor is the receipt of an incentive award usually contingent on teacher effectiveness.
13. For more information visit, <http://cppe.wceruw.org/index.php>
14. In his review of the different types of corruption and their potential causes in the education sector, Heyneman (2004) recommends four types of reform to minimise the risk. These include educational structures, the processes of management and adjudication, the mechanism of prevention when wrongdoing occurs, and the system of sanctions.
15. Carnoy, Brodziak, Molina, and Solcias (2007) report on constraints of a school-based incentive programme when using student performance data to reward performance on inter-cohort gains as opposed to intra-cohort test score gains.
16. For a more complete description of issues related to VAM see Chapter 5 by Dan Goldhaber.
17. Ballou (2009) discussed the implications of scale properties in value-added measurements. He analysed several possible methods of formulating value-added estimates from tests that are scored according to Item Response Theory (IRT). He found measuring ability on an interval scale using IRT requires a conjoint structure between students and test items, conditions that are both difficult to meet and verify. Value-added assessments of teacher quality may introduce volatility due to the scale properties of tests that may incorrectly be attributed to teacher effectiveness.
18. Jacob and Levitt's (2003) analysis of cheating within the **Chicago** Public School system provided two methods of detection. The first method involved searching for large-scale uniform patterns in student responses, patterns which would be statistically impossible in normal student responses. Secondly, unusually high student performances in one year that are followed by declines in performance in subsequent years were selected for closer observation. In international programs in **Israel**, **Kenya**, and **India** there were penalties for students who did not take the exam in an effort to prevent schools and teachers from intentionally discouraging low-performing students from test participation. In each of these cases, test participation rates either stayed the same or significantly increased (Lavy, 2004; Glewwe, Ilias, and Kremer, 2003; Muralidharan and Sundararaman, 2008).
19. Table 3.2 builds on evidence reported in Taylor, Springer, and Ehlert (2009).
20. The information displayed in Figure 3.3 does not include those teachers that participated in the TEEG program but did not earn a bonus award during the 2006-07 school year.
21. The **Houston** Chronicle posts information on bonus award amounts delineated by employee name, position, and school (see, for example, <http://www.chron.com/news/houston/bonuses>). The **Austin** Statesman similarly publishes names of teachers earning bonuses under the district's Strategic Compensation Initiative.
22. Burns, Gardner, Muesswen (2009) provide additional discussion within the context of **Austin** Independent School District's Strategic Compensation Initiative.
23. In a related study, Ramamoorthy and Carroll (1998) reported a significant relationship between high levels of individualism and alternative human resource management practices such as merit pay.

References

- Adnett, N.** (2003), "Reforming Teachers' Pay: Incentive Payments, Collegiate Ethos and UK Policy", *Cambridge Journal of Economics*, Vol. 27, No. 1, pp. 145-147.
- Asch, B.** (1990), Navy Recruiter Productivity and the Freeman Plan, *RAND Monograph R-3713-FMP*, Santa Monica, CA.
- Asch, B.** (2005), "The Economic Complexities of Incentive Reforms", in R. Klitgaard and P.C. Light (eds.), *High-Performing Government: Structure, Leadership, Incentives*, RAND Corporation, Santa Monica, CA.
- Atkinson, A, S. Burgess, B. Croxson, P. Gregg, C. Propper, H. Slater and D. Wilson** (2009), "Evaluating the Impact of Performance-Related Pay for Teachers in England", *Labour Economics*, Vol. 16, No. 3, pp. 251-261.
- Baker, G.** (1992), "Incentive Contracts and Performance Measurement", *Journal of Political Economy*, Vol. 100, No. 3, pp. 598-614.
- Baker, G.** (2002), "Distortion and Risk in Optimal Performance Contracts", *The Journal of Human Resources*, Vol. 37, No. 4, pp. 728-751.
- Baker, G., M. Jensen and K. Murphy** (1988), "Compensation and Incentives: Practice vs. Theory", *Journal of Finance*, Vol. 43, No. 3, 593-616.
- Ballou, D.** (2001), "Pay for Performance in Public and Private Schools", *Economics of Education Review*, Vol. 20, No. 1, pp. 51-61.
- Ballou, D.** and **M. Podgursky** (1997), "Reforming Teacher Training & Recruitment: A Critical Appraisal of the Recommendations of the National Commission on Teaching and America's Future", *Government Union Review*, Vol. 17, No. 4, pp. 1-53.
- Ballou, D.** and **M. Podgursky** (2009), "Responses to No Child Left Behind Among Traditional Public Schools", *Peabody College of Vanderbilt University Working Paper*, Nashville, TN.
- Banker, R.** and **S. Datar** (1989), "Sensitivity, Precision, and Linear Aggregation of Signals for Performance Evaluation", *Journal of Accounting Research*, Vol. 27, No. 1, pp. 21-39.
- Baron, J.** and **D. Kreps** (1999), *Strategic Human Resources: Frameworks for General Managers*, Wiley, New York.
- Berg, P., E. Appelbaum, T. Bailey and A. Kalleberg** (1996), "The Performance Effects of Modular Production in the Apparel Industry", *Industrial Relations*, Vol. 35, No. 3, pp. 356-373.
- Bishop, J.** (1996). "Signaling, Incentives, and School Organization in France, the Netherlands, Britain, and the United States", in E. Hanushek and D. Jorgenson (eds.), *Improving America's Schools: The Role of Incentives*, National Research Council, Washington, D.C.
- Bok, D.** (1993), *The Cost of Talent: How Executives and Professionals are Paid and How It Affects America*, The Free Press, New York.
- Booher-Jennings, J.** (2005), "Below the Bubble: 'Educational Triage' and the Texas Accountability System", *American Educational Research Journal*, Vol. 42, No. 2, pp. 231-268.
- Bourne, R.** and **B. MacArthur** (1970), *The Struggle for Education: 1870-1970*, School Masters Publishing Company, London.
- Bowles, S.** and **H. Gintis** (2002), "Schooling in Capitalist America Revisited", *Sociology of Education*, Vol. 75, No. 1, pp. 1-18.
- Bryk, A.** and **B. Schneider** (2002), *Trust in Schools: A Core Resource for Improvement*, Russell Sage Foundation, New York.
- Burgess, S., B. Croxson, P. Gregg and C. Propper** (2001), "The Intricacies of the Relationship Between Pay and Performance for Teachers: Do Teachers Respond to Performance Related Pay Schemes", *CMPO Working Paper*, Bristol.

- Chamberlin, R., T. Wragg, G. Haynes, and C. Wragg** (2002), "Performance-Related Pay and the Teaching Profession: A Review of the Literature", *Research Papers in Education*, Vol. 17, No. 1, pp. 31-49.
- Che, Y. and S. Yoo** (2001), "Optimal Incentives for Teams", *The American Economic Review*, Vol. 91, No. 3, pp. 525-541.
- Chiang, H.** (forthcoming), "How Accountability Pressure of Failing Schools Affects Student Achievement", *Journal of Public Economics*.
- Chiang, F. and T. Birtch** (2005), "A Taxonomy of Reward Preferences: Examining Country Differences", *Journal of International Management*, Vol. 11, pp. 357-375.
- Clotfelter, C., H. Ladd and J. Vigdor** (2004), "Do School Accountability Systems Make It More Difficult for Low-performing Schools to Attract and Retain High-Quality Teachers?", *Journal of Policy Analysis and Management*, Vol. 23, No. 2, pp. 251-271.
- Clotfelter, C., H. Ladd and J. Vigdor** (2007), "Are Teacher Absences Worth Worrying About in the U.S.?", *NBER Working Paper 13648*, National Bureau of Economic Research, Cambridge, MA.
- Corcoran, S., W. Evans and R. Schwab** (2004), "Women, the Labor Market, and the Declining Relative Quality of Teachers", *Journal of Policy Analysis and Management*, Vol. 23, No. 3, pp. 449-470.
- Courty, P. and G. Marschke** (1997), "Measuring Government Performance: Lessons from a Federal Job-Training Program", *American Economic Review*, Vol. 87, No. 2, pp. 383-388.
- Courty, P. and G. Marschke** (2003), "Dynamics of Performance-Measurement Systems", *Oxford Review of Economic Policy*, Vol. 19, pp. 68-284.
- Courty, P., C. Heinrich and G. Marschke** (2005), "Setting the Standard in Performance Measurement Systems", *International Public Management Journal*, Vol. 8, No. 3, pp. 321-347.
- Crewson, E.** (1997), "Public-Service Motivation: Building Empirical Evidence of Incidence and Effect", *Journal of Public Administration Research and Theory*, Vol. 7, No. 4, pp. 499-518.
- Cutler, T. and B. Waine** (2004), "Performance Without Pay? Managing School Budgets Under Performance Related Pay: Evidence From the First Year of PRP", *Financial Accountability and Management*, Vol. 20, No. 1, pp. 57-75.
- Danielson, C.** (1996), *Enhancing Professional Practice: A Framework for Teaching*. Association for Supervision and Curriculum Development, Alexandria, VA.
- Deutsch, M.** (1985), *Distributive Justice: A Social-Psychological Perspective*, Yale University Press, New Haven, CT.
- Dixit, A.** (2002), "Incentives and Organizations in the Public Sector", *Journal of Human Resources*, Vol. 37, No. 4, pp. 696-727.
- Duflo, E., R. Hanna and S. Ryan** (2005), "Monitoring Works: Getting Teachers to Come to School", *NBER Working Paper 11880*, National Bureau of Economic Research, Cambridge, MA.
- Eberts, R.** (2007), "Teacher Unions and Student Performance: Help or Hindrance?", *The Future of Children*, Vol. 17, No. 1, pp. 175-200.
- Eberts, R., K. Hollenbeck and J. Stone** (2002), "Teacher Performance Incentives and Student Outcomes", *Journal of Human Resources*, Vol. 37, No. 4, 913-927.
- Ehrenberg, R. and R. Smith** (1994), *Modern Labor Economics: Theory and Public Policy*, Fifth Edition, Harper Collins.
- Elmore, R.F., C.H. Abelman, and S.H. Fuhrman**, (1996). "The new accountability in state education reform: From process to performance". In H. Ladd (ed.), *Holding schools accountable: Performance-based reform in education*. (pp. 65±98). Washington, DC: Brookings Institution.
- Feltham, G. and J. Xie** (1994), "Performance Measure Congruity and Diversity in Multi-Task Principal/Agent Relations", *The Accounting Review* Vol. 69, No. 3, pp. 429-453.
- Figlio, D. and C. Rouse** (2007), "Do Accountability and Voucher Threats Improve Low-Performing Schools", *Journal of Public Economics*, Vol. 90, No. 1-2, pp. 239-255.
- Firestone, W.** (1994), "Redesigning Teacher Salary Systems for Educational Reform", *American Educational Research Journal*, Vol. 31, No. 3, pp. 549-574.

- Freeman, R.** and **A. Gelber** (2006), "Optimal Inequality/Optimal Incentive: Evidence from a Tournament", *NBER Working Paper 12588*, National Bureau of Economic Research, Cambridge, MA.
- Freixas, X., R. Guesnerie** and **J. Tirole** (1985), "Planning under Incomplete Information and the Ratchet Effect", *Review of Economic Studies*, Vol. 52, No. 2, pp. 173-191.
- Gaynor, M.** and **M. Pauly** (1990), "Compensation and Productive Efficiency in Partnerships: Evidence from Medical Group Practice", *Journal of Political Economy*, Vol. 98, No. 3, pp. 543-577.
- Gerhart, B.** and **G. Milkovich** (1990), "Organizational Differences in Managerial Compensation and Financial Performance", *Academy of Managerial Journal*, Vol.33, No. 4, pp. 663-691.
- Ginther, D., R. Haveman** and **B. Wolfe** (2000), "Neighborhood Attributes as Determinants of Children's Outcomes: How Robust are the Relationships", *The Journal of Human Resources*, Vol. 35, No. 4, pp. 603-642.
- Glewwe, P., A. Holla** and **M. Kremer** (2009), "Teacher Incentives in the Developing World", in M.G. Springer (ed.), *Performance Incentives: Their Growing Impact on American K-12 Education*, Brookings Institution Press, Washington, D.C.
- Glewwe, P., N. Ilias** and **M. Kremer** (2008), *Teacher Incentives in the Developing World*, Harvard University, Cambridge, MA.
- Goldhaber, D.** (2002), "The Mystery of Good Teaching," *Education Next*, Vol. 2, No. 1.
- Goldhaber, D.** (2009), "The Politics of Teacher Pay Reform", in Springer, M.G. (ed.), *Performance Incentives: Their Growing Impact on American K-12 Education*, Brookings Institution Press, Washington, D.C.
- Goldhaber, D., D. Brewer** and **D. Anderson** (1999), "A Three-Way Error Components Analysis of Educational Productivity", *Education Economics*, Vol. 7, No. 3, pp. 191-208.
- Gonring, P., P. Teske** and **B. Jupp** (2007), *Pay-for-Performance Teacher Compensation: An Inside View of Denver's ProComp Plan*, Harvard Education Press, Cambridge, MA.
- Gramlich, E.** and **P. Koshel** (1975), *Educational Performance Contracting*, Brookings Institution, Washington, D.C.
- Guthrie, J.** and **C. Prince** (2009), "Paying for and Sustaining a Performance-Based Compensation System", Center for Educator Compensation Reform, Washington, D.C.
- Hallinger, P.** and **J. Murphy** (1986). "The Social Context of Effective Schools", *American Journal of Education*, Vol. 94, No. 3, pp. 328-355.
- Hamilton, L.** (2005), "Lessons from Performance Measurement in Education", in R. Klitgaard and P.C. Light (eds.), *High-Performing Government: Structure, Leadership, Incentives*, RAND Corporation, Santa Monica, CA.
- Hamilton, B., J. Nickerson** and **H. Owan** (2003), "Team Incentives and Worker Heterogeneity: An Empirical Analysis of the Impact of Teams on Productivity and Participation", *Journal of Political Economy*, Vol. 111, No. 3, pp. 465-497.
- Hannaway, J.** (1992), "Higher Order Skills, Job Design, and Incentives: An Analysis and Proposal", *American Educational Research Journal*, Vol. 29, No. 1, pp. 3-21.
- Hannaway, J.** (1996). "Management Decentralization and Performance-Based Incentives: Theoretical Considerations for Schools", in E. Hanushek and D. Jorgenson (eds.), *Improving America's Schools: The Role of Incentives*, National Research Council, Washington, D.C.
- Hansen, D.** (1997), "Worker Performance and Group Incentives: A Case Study," *Industrial Labor Relations Review*, Vol. 51, pp. 37-49.
- Heinrich, C.J.** (2007), "Evidence-Based Policy and Performance Management: Challenges and Prospects in Two Parallel Movements", *The American Review of Public Administration*, Vol. 37, No. 3, pp. 255-277.
- Heneman, H., A. Milanowski** and **S. Kimball** (2007), "Teacher Performance Pay: Synthesis of Plans, Research, and Guidelines for Practice", *CPRE Policy Briefs RB-46*, Madison, WI.
- Hofstede, G.** (1980), *Culture's Consequences: International Differences in Work-Related Values*, Sage Publications, Beverly Hills, CA/ London.

- Hofstede, G.** (2001), *Culture's Consequences: Comparing Values, Behaviors, Institutions, and Organizations Across Nations*, Sage Publications, Thousand Oaks, CA.
- Hollensbe, E.** and **J. Guthrie** (2000), "Group Pay-for-Performance Plans: The Role of Spontaneous Goal Setting", *The Academy of Management Review*, Vol. 25, No. 4, pp. 864-872.
- Holmstrom, B.** (1979). "Moral Hazard and Observability", *The Bell Journal of Economics*, Vol. 10, No. 1, pp. 74-91.
- Holmstrom, B.** (1982), "Moral Hazard in Teams", *Bell Journal of Economics*, Vol. 13, No. 2, pp. 324-340.
- Holmstrom, B.** and **P. Milgrom** (1987), "Aggregation and Linearity in the Provision of Intertemporal Incentives", *Econometrica*, Vol. 55, No. 2, pp. 303-328.
- Holmstrom, B.** and **P. Milgrom** (1991), "Multitask Principal-Agent Analysis: Incentive Contracts, Asset Ownership, and Job Design", *Journal of Law, Economics, and Organization*, Vol. 7, pp. 24-52.
- Hoxby, C.** and **A. Leigh** (2004), "Pulled Away or Pushed Out? Explaining the Decline of Teacher Aptitude in the United States", *American Economic Review*, Vol. 94, No. 2, pp. 236-240.
- Neale, M., V. Huber** and **G. Northcraft** (1987), "The Framing of Negotiations: Contextual Versus Task Frames", *Organizational Behavior and Human Decision Processes*, Vol. 39, pp. 228-241.
- Indjejikian, R., P. Lenk** and **D. Nanda** (2000), "Targets, Standards and Performance Expectations: Evidence from Annual Bonus Plans", *SSRN Working Paper*, http://papers.ssrn.com/sol3/papers.cfm?abstract_id=213628, accessed September 20, 2009.
- Jacob, B.** and **S. Levitt** (2003), "Rotten Apples: An Investigation of the Prevalence and Predictors of Teacher heating", *Quarterly Journal of Economics*, Vol. 118, No. 3, pp. 843-877.
- Jacob, B.** and **L. Lefgren** (2005), "Principals as Agents: Subjective Performance Measures in Education", *KSG Working Paper No. RWP05-040*, Cambridge, MA.
- Jacobson, S.** (1992), "Performance-Related Pay for the Teacher: The American Experience", in H. Tomlinson (ed.), *Performance-Related Pay in Education*, Routledge, London.
- Johnson, S.M.** (2004). "Paralysis or Possibility: What Do Teacher Unions and Collective Bargaining Bring", in R. Henderson, W. Urban, and P. Wolman (eds.), *Teacher Unions and Education Policy: Retrenchment or Reform?*, Elsevier, Oxford, U.K.
- Juerges, H., W. Richter** and **K. Schneider** (2004), "Teacher Quality and Incentives: Theoretical and Empirical Effects of Standards on Teacher Quality", *CESifo Working Paper No. 1296*, CESifo, Munich.
- Kandel, E.** and **E. Lazear** (1992), "Peer Pressure and Partnerships", *The Journal of Political Economy*, Vol. 100, No. 4, pp. 801-817.
- Kane, T.** and **D. Staiger** (2001), "Improving School Accountability Measures", *NBER Working Paper 8156*, National Bureau of Economic Research, Cambridge, MA.
- Kelley, C.** (1999), "The Motivational Effects of School-Based Performance Awards", *Journal of Personnel Evaluation in Education*, Vol. 12, No. 4, pp. 54-78.
- Kelley, C.** (2002), "Financial Incentives in State Accountability Systems: Performance Pay for Teachers", Center on Reinventing Public Education and the Washington State Academic Achievement and Accountability Commission, Seattle, WA.
- Kimball, S., B. White, A. Milanowski** and **G. Borman** (2004), "Examining the Relationship Between Teacher Evaluation and Student Assessment Results in Washoe County", *Peabody Journal of Education*, Vol. 79, No. 4, pp. 54-78.
- Kingdon, G.** and **F. Teal** (2008), "Teacher Unions, Teacher Pay and Student Achievement in India", *CESifo Working Paper 2428*, CESifo, Munich.
- Kohn, A.** (1999), *Punished by Rewards: The Trouble with Gold Stars, Incentive Plans, A's, Praise, and Other Bribes*, Houghton Mifflin, Boston, MA.
- Kole, S.** (1997), "The Complexity of Compensation Contracts", *Journal of Financial Economics*, Vol. 43, pp. 79-104.

- Kremer, M.** and **D. Chen** (2001), "Interim Report on a Teacher Incentive Program in Kenya", University of Chicago, Chicago, IL <http://home.uchicago.edu/~dlc/papers/Preschool.pdf>.
- Ladd, H.** (1999), "The Dallas School Accountability and Incentive Program: An Evaluation of Its Impacts on Student Outcomes", *Economics of Education Review*, Vol. 18, No. 1, pp. 1-16.
- Lavy, V.** (2002), "Evaluating the Effect of Teachers' Group Performance Incentives on Pupil Achievement", *Journal of Political Economy*, Vol. 110, No. 6, pp. 1286-1317.
- Lavy, V.** (2004), "Performance Pay and Teachers' Effort, Productivity and Grading Ethics", *NBER Working Paper 10622*, National Bureau for Economic Research, Cambridge, MA.
- Lavy, V.** (2007), "Using Performance-Based Pay to Improve the Quality of Teachers", *The Future of Children*, Vol. 17, No. 1, pp. 87-109.
- Lavy, V.** (2008), "Gender Differences in Market Competitiveness in a Real Workplace: Evidence from Performance-Based Tournaments Among Teachers", *NBER Working Paper 14338*, National Bureau of Economic Research, Cambridge, MA.
- Lazear, E.** (1995), *Personnel Economics*, MIT Press, Cambridge, MA.
- Lazear, E.** (1998), *Personnel Economics for Managers*, Wiley, New York.
- Lazear, E.** (2000), "Performance Pay and Productivity", *American Economic Review*, Vol. 90, No. 5, 1346-1361.
- Lazear, E.** (2003), "Teacher Incentives", *Swedish Economic Policy Review*, Vol. 10, pp. 197-213.
- Lazear, E.** and **M. Gibbs** (2009), *Personnel Economics in Practice*, Wiley, Hoboken, NJ.
- Leventhal, T.** and **J. Brooks-Gunn** (2004), "A Randomized Study of Neighborhood Effects on Low-Income Children's Educational Outcomes", *Developmental Psychology*, Vol. 40, No. 4, pp. 488-507.
- Lewis, J.** and **M. Springer** (2008), "Performance Incentives in Texas: Why Schools Chose Not to Participate", *Working Paper #2008-18*, National Center on Performance Incentives, Nashville, TN.
- Liang, X.** (1999), "Teacher Pay in 12 Latin American Countries: How Does Teacher Pay Compare to Other Professions, What Determines Teacher Pay, and Who Are the Teachers?", *LCSHD Paper Series No. 49*, World Bank, Washington, D.C.
- Locke, E.A.** (1968), "Toward a Theory of Task Motivation and Incentives", *Organizational Behavior and Human Performance*, Vol. 3, pp. 157-189.
- Locke, E.** and **G. Latham** (1990), *A Theory of Goal Setting and Task Performance*, Prentice-Hall, Englewood Cliffs, NJ.
- Locke, E., G. Latham** and **K. Smith** (1990). *A Theory of Goal Setting and Task Performance*, Prentice-Hall, Upper Saddle River, NJ.
- Locke, E., L. Saari, K. Shaw, G. Latham** (1981), "Goal Setting and Task Performance: 1968-1980", *Psychological Bulletin*, Vol. 90, No. 1, pp. 125-152.
- Longenecker, C., D. Gioia** and **H. Sims** (1987), "Behind the Mask: The Politics of Employee Appraisal", *Academy of Management Executive*, Vol. 1, No. 3, pp. 183-193.
- Lortie, D.** (1979), *Schoolteacher: A Sociological Study*, University of Chicago Press, Chicago, IL.
- Malen, B. (1999), "On Rewards, Punishments, and Possibilities: Teacher Compensation as an Instrument for Education Reform", *Journal of Personnel Evaluation in Education*, Vol. 12, No. 4, pp. 387-394.
- Marsden, D.** and **R. Belfield** (2006), "Pay for Performance Where Output is Hard to Measure: The Case of Performance Pay for School Teachers", in B. Kaufman, E. Bruce and D. Lewin (eds.), *Advances in Industrial and Labor Relations*, JAI Press, London.
- McAdams, J.** and **E. Hawk** (1992), *Capitalizing on Human Assets*, American Compensation Association, Scottsdale, AZ.
- McCaffrey, D., B. Han** and **J. Lockwood** (2009), "Incentive System Design and Measurement", in M.G. Springer (ed.), *Performance Incentives: Their Growing Impact on American K-12 Education*, Brookings Institution Press, Washington, D.C.
- McCullum, S.** (2001), "How Merit Pay Improves Education", *Educational Leadership*, Vol. 58, No. 5, pp. 21-24.

- McEwan, P.** (1999), "Recruitment of Rural Teachers in Developing Countries: An Economic Analysis", *Teaching and Teacher Education*, Vol. 15, No. 8, pp. 849-859.
- Merchant, K.A.** and **J. Manzoni** (1989), "The Achievability of Budget Targets in Profit Centers: A Field Study", *The Accounting Review*, Vol. 64, pp. 539-558.
- Milgrom, P.** and **J. Roberts** (1990), "Rationalizability, Learning, and Equilibrium in Games with Strategic Complementarities", *Econometrica*, Vol. 58, No. 6, pp. 1255-1277.
- Milgrom, P.** and **J. Roberts** (1992), *Economics, Organization & Management*, Prentice Hall, Englewood Cliffs, CA.
- Milanowski, A.** (2003), "The Varieties of Knowledge and Skill-Based Pay Design: A Comparison of Seven New Pay Systems for K-12 Teachers", *Education Policy Analysis Archives*, Vol. 11, No. 4, <http://epaa.asu.edu/epaa/v11n4/>, accessed April 17, 2009.
- Milanowski, A.** (2004), "The Relationship between Teacher Performance Evaluation Scores and Student Achievement: Evidence from Cincinnati", *Peabody Journal of Education*, Vol. 79, No. 4, pp. 33-53.
- Mirrlees, J.** (1999), "The Theory of Moral Hazard and Unobservable Behaviour: Part I", *Review of Economic Studies*, Vol. 66, No. 1, pp. 3-21.
- Mizala, A.** and **P. Romaguera** (2003), *Scholastic Performance and Performance Awards*, Centro de Economica Aplicada, Santiago, Chile.
- Mizala, A.** and **P. Romaguera** (2004), "Schools and Teachers Performance Incentives: The Latin American Experience", *International Journal of Educational Development*, Vol. 24, No. 6, pp. 739-754.
- Mullen, C.** and **K. Slagle** (2007), "Teacher Perceptions of Systems-Driven Quality Teaching Initiatives", *Academic Leadership*, Vol. 3, No. 1, pp. 1-20.
- Mullen, K., R. Frank** and **M. Rosenthal** (2009), "Can You Get What You Pay For? Pay-for-Performance and the Quality of Healthcare Providers", *NBER Working Paper 14886*, National Bureau of Economic Research, Cambridge, MA.
- Muralidharan, K.** and **V. Sundararaman** (2008), "Teacher Incentives in Developing Countries: Experimental Evidence from India", Harvard University, Cambridge, MA.
- Murnane, R.** and **D. Cohen** (1985), "*'Merit Pay' and the Nature of the Employment Relationship: Lessons from Public School Systems*. Education Resources Information Center, Washington, D.C., http://www.eric.ed.gov/ERICDocs/data/ericdocs2sql/content_storage_01/0000019b/80/2f/4e/c7.pdf.
- Murnane, R.** and **D. Cohen** (1986), "Merit Pay and the Evaluation Problem: Why Most Merit Pay Plans Fail and Few Survive", *Harvard Education Review*, Vol. 56, pp. 1-17.
- Murphy, K.** (1999), "Executive Compensation", in *Handbook of Labor Economics, Vol. III*, Ashenfelter, O. and D. Card (eds.), North Holland, Oxford.
- Murphy, K.** (2001), "Performance Standards in Incentive Contracts", *Journal of Accounting and Economics*, Vol. 30, pp. 245-278.
- Murphy, T.** and **J. Margulies** (2004), "Performance Appraisals". Paper presented at ABA Employment and Law Section Conference, Equal Opportunity Employment Committee, Needham Heights, MA, March 24-27.
- Nalbantian, H.** and **A. Schotter** (1997), "Productivity Under Group Incentives: An Experimental Study", *The American Economic Review*, Vol. 87, No. 3, pp. 314-341.
- Neal, D.** (2009), "Designing Incentive Systems for Schools", in M.G. Springer (ed.), *Performance Incentives: Their Growing Impact on American K-12 Education*, Brookings Institution Press, Washington, D.C.
- Neal, D.** and **D. Schanzenbach** (forthcoming), "Left Behind By Design: Proficiency Counts and Test-Based Accountability", *Review of Economics and Statistics*.
- Odden, A.** and **C. Kelley** (1997), *Paying Teachers for What They Know and Do: New and Smarter Compensation Strategies to Improve Schools*, Corwin Press, Thousand Oaks, CA.

- Odden, A., G. Borman, and M. Fermanich** (2004), "Assessing Teacher, Classroom, and School Effects, Including Fiscal Effects", *Peabody Journal of Education*, Vol. 79, No. 4, pp. 4-32.
- Oyer, P.** (1998), "Fiscal Year Ends and Nonlinear Incentive Contracts: The Effect on Business Seasonality", *The Quarterly Journal of Economics*, Vol. 113, No. 1, pp. 149-185.
- Paarsch, H. and B. Shearer** (2000), "Piece Rates, Fixed Wages, and Incentive Effects: Statistical Evidence from Payroll Records", *International Economic Review*, Vol. 41, No. 1, pp. 59-92.
- Perry, J., T. Engbers and S. Jun** (2009), "Back to the Future? Performance-Related Pay, Empirical Research, and the Perils of Persistence", *Public Administration Review*, Vol. 69, No. 1, pp. 39-51.
- Perry, J. and L. Porter** (1982), "Factors Affecting the Context for Motivation in Public Organizations", *The Academy of Management Review*, Vol. 7, No. 1, pp. 89-98.
- Pfeffer, J.** (1995), *Competitive Advantage Through People: Unleashing the Power of the Work Force*, Harvard Business School Press, Boston, MA.
- Pfeffer, J. and N. Langston** (1993), "The Effect of Wage Dispersion on Satisfaction, Productivity, and Working Collaboratively: Evidence from College and University Faculty", *Administrative Science Quarterly*, Vol. 38, No. 2, pp. 382-407.
- Podgursky, M. and M. Springer** (2007), "Teacher Performance Pay: A Review", *Journal of Policy Analysis and Management*, Vol. 26, No. 4, pp. 909-949.
- Prendergast, C.** (1999), "The Provision of Incentives in Firms", *Journal of Economic Literature*, Vol. 37, No.1, pp. 7-63.
- Prendergast, C. and R. Topel** (1993), "Discretion and Bias in Performance Evaluation", *European Economic Review*, Vol. 37, No. 2-3, pp. 909-950.
- Rainey, H.** (1982), "Reward Preferences Among Public and Private Managers: In Search of the Service Ethic", *The American Review of Public Administration*, Vol. 16, No. 4, pp. 288-302.
- Rapple, B.** (1990). Payment by Educational Results: An Idea Whose Time Has Gone? Thirty-Five Years of Experimentation with Education Efficiency in England (1862-1897), *MacArthur-Spencer Series Number 14*, Chicago, IL.
- Reback, R.** (2008), "Teaching to the Rating: School Accountability and the Distribution of Student Achievement", *Journal of Public Economics*, Vol. 92, No. 5-6, pp. 1394-1415.
- Richardson, R.** (1999), "Performance Related Pay in Schools: An Assessment of The Green Papers", London School of Economics and Political Science, London.
- Rivkin, S., E. Hanushek and J. Kain** (2005), "Teachers, Schools, and Academic Achievement", *Econometrica*, Vol. 73, No. 2, pp. 417-458.
- Rosen, S.** (1986), "The Theory of Equalizing Differences", in O. Ashenfelter and R. Layard (eds.), *Handbook of Labor Economics*, Vol. 1, North-Holland, Oxford.
- Rosenholtz, S.** (1984), "Myths: Political Myths About Reforming Teaching", National Institute of Education, Washington, D.C.
- Rothstein, R.** (2009), "How Scholarship and Experience in Other Fields Inform Exploration of Performance Incentives in Education", in M.G. Springer (ed.), *Performance Incentives: Their Growing Impact on American K-12 Education*, Brookings Institution Press, Washington, D.C.
- Schacter, J. and Y. Thum** (2004), "Paying for High- and Low-Quality Teaching", *Economics of Education Review*, Vol. 23, No. 4, pp. 411-430.
- Schmitt, N., R. Noe and R. Gottschalk** (1986), "Using the Lens Model to Magnify Raters' Consistency, Matching, and Shared Bias", *Academy of Management Journal*, Vol. 29, pp. 130-139.
- Schneider, M., P. Teske and M. Marshall** (2000), *Choosing Schools: Consumer Choice and the Quality of American Schools*, Princeton University Press, Princeton, NJ.
- Skinner, B.** (1981), "Selection by Consequences", *Science*, Vol. 213, No. 4507, pp. 501-504.

- Springer, M.** (2008), "The Influence of an NCLB Accountability Plan on the Distribution of Student Test Score Gains." *Economics of Education Review*, Vol. 27, No. 5, pp. 556-563.
- Springer, M.** (2009), "Rethinking Teacher Compensation Policies: Why Now, Why Again?", in M.G. Springer (ed.), *Performance Incentives: Their Growing Impact on American K-12 Education*, Brookings Institution Press, Washington, D.C.
- Springer, M., D. Ballou and A. Peng** (2008), *Impact of the Teacher Advancement Program on Student Test Score Gains: Findings from an Independent Appraisal*, National Center on Performance Incentives, Nashville, TN.
- Springer, M., M. Podgursky, J. Lewis, M. Ehlert, T. Gronberg, L. Hamilton, D. Jansen, O. Lopez, A. Peng, B. Stecher and L. Taylor** (2008), *Texas Educator Excellence Grant (TEEG), Program: Year Two Evaluation Report*, National Center on Performance Incentives, Nashville, TN.
- Springer, M., M. Podgursky, J. Lewis, M. Ehlert, T. Gronberg, D. Jansen, O. Lopez, A. Peng, B. Stecher and L. Taylor** (2009), *Governor's Educator Excellence Grant (GEEG), Program: Year Two Evaluation Report*, National Center on Performance Incentives, Nashville, TN.
- Springer, M. and M. Winters** (2009), *New York City's School-Wide Bonus Pay Program: Early Evidence From a Randomized Trial*, National Center on Performance Incentives, Nashville, TN.
- Suen, H. and L. Yu** (2006), "Chronic Consequences of High-Stakes Testing? Lessons from the Chinese Civil Service Exam", *Comparative Education Review*, Vol. 50, No. 1, pp. 46-65.
- Taylor, L. and M. Springer** (2009), "Optimal Incentives for Public Sector Workers: The Case of Teacher-Designed Incentive Pay in Texas", *National Center on Performance Incentives Working Paper*, Nashville, TN.
- Taylor, L., M. Springer and M. Ehlert** (2009), "Characteristics and Determinants of Teacher-Designed Pay for Performance Plans: Evidence From Texas' Governor's Educator Excellence Grant Program", in M.G. Springer (ed.), *Performance Incentives: Their Growing Impact on American K-12 Education*, Brookings Institution Press, Washington, D.C.
- Tomlinson, H.** (2000), "Proposals for Performance Related Pay in English Schools", *School Leadership and Management*, Vol. 20, No. 3, pp. 281-298.
- Vroom, V.** (1964), *Work and Motivation*, Wiley, New York, NY.
- Weiss, A.** (1987), "Incentives and Worker Behavior: Some Evidence", in H. Nalbantian (ed.), *Incentives, Cooperation and Risk Sharing*, Rowan & Littlefield, Latham, MD.
- Weisberg, D., S. Sexton, J. Mulhern and D. Keeling** (2009), *The Widget Effect: Our National Failure to Acknowledge and Act on Differences in Teacher Effectiveness*, The New Teacher Project, Brooklyn, NY.
- Weitzman, M.** (1980), "The Ratchet Principle and Performance Incentives", *Bell Journal of Economics*, Vol. 11, No. 1, pp. 302-308.
- West, M. and P. Peterson** (2006), "The Efficacy of Choice Threats Within School Accountability Systems: Results from Legislatively Induced Experiments", *The Economic Journal*, Vol. 116, No. 510, pp. C46-C62.
- Wilson, W.** (1987), *The Truly Disadvantaged: The Inner City, the Underclass, and Public Policy*, University of Chicago Press, Chicago, IL.
- Wise, S.** (2004), "Work-life Balance and Careers in NHS Nursing and Midwifery", Employment Research Institute, Napier University, Edinburgh.
- Zenger, T.** (1992), "Why Do Employers Only Reward Extreme Performance? Examining the Relationship Among Performance, Pay, and Turnover", *Administrative Science Quarterly*, Vol. 37, No. 2, pp. 198-219.