Group Incentives for Teachers and their Effect on Student Learning: A Systematic Review of Theory and Evidence

N. Tirivayi, W. Groot and H. Maassen van den Brink
Group Incentives for Teachers and their Effect on Student Learning: A Systematic Review of Theory and Evidence

Nyasha Tirivayi†, Henriette Maasen van den Brink‡ and Wim Groot§

Abstract

This review examines the effects of teachers’ group incentives on student achievement. This paper adds value to literature by first assessing theoretical arguments and then empirical studies from both OECD and developing countries published between 1990 and 2011. Studies from developing countries reported positive effects of group incentives on student test scores. However, all experimental studies conducted in developed countries reported insignificant effects, unlike non-experimental studies which reported positive effects. It is difficult to ascertain whether the key to successful group incentives in teaching emanates from the size of the incentive, group size, teacher intrinsic motivation or type of incentive (rank type vs. non-rank type). We conclude that current nascent body of empirical evidence has unclear policy implications on the long term learning gains from group incentives and there is need for additional experimental research to establish firm conclusions and policy recommendations.

Keywords: group incentives, teachers, student achievement, performance pay, education

†Corresponding author. Top Institute for Evidence Based Education Research (TIER), Maastricht University, P.O Box 6200 MD, The Netherlands. nyasha.tirivayi@maastrichtuniversity.nl
‡Top Institute for Evidence Based Education Research (TIER), Maastricht University, P.O Box 6200 MD, The Netherlands
§Top Institute for Evidence Based Education Research (TIER), Maastricht University, P.O Box 6200 MD, The Netherlands
Introduction

Education plays a key role in human capital development, labour participation and economic growth. Hence there has always been an intense focus on improving the quality of schools. Research has already established that current teacher salaries based on qualifications are poor predictors of student achievement (Muralidharan & Sundararaman, 2011; Rivkin, Hanushek, & Kain, 2005; Gordon, Kane, & Staiger, 2006). Recently there has been growing attention by policymakers and researchers towards improving teacher performance and teacher quality in schools. One of the ideas attracting wide interest is incentive pay or performance based incentives. Performance based incentives have their intellectual roots in the private sector literature (Oyebolu & Muraina, 2010). Competition in the private sector spurred the development of strategies such as performance based incentives to maximize productivity and efficiency. In the field of education, advocates of performance incentives, regard large firms that implement rewards as models for teaching and argue that they have similar environments to teaching (Oyebolu & Muraina, 2010; Ballou & Podgursky, 2001; Odden, 2000). Advocates also argue that performance pay in education is cost-efficient as you do not reward everyone but expect everyone to increase effort and that it improves the governance of schools by increasing the efficiency of resource allocation (Harvey-Beavis, 2003).

Performance-based incentives in education initially emerged in the United States. They developed mostly in response to the criticism of single salary schedule or uniform pay which was first implemented in Colorado and Iowa in 1921 (Podgursky & Springer, 2007). The major critique being that uniform pay would standardize remuneration and hinder the adjustment of a teacher’s pay to reflect performance and labor market reality (Podgursky & Springer, 2007). In the last two decades several countries in the OECD and now in the developing world have adopted performance based strategies in education (Harvey-Beavis, 2003). Examples of OECD countries include Australia, Belgium, France, Mexico, New
Zealand and the United Kingdom, while India and Kenya are some of the developing countries introducing performance pay in education (Harvey-Beavis, 2003; Muralidharan & Sundararaman, 2011). Some performance based incentives for teachers are based on added bonus to a normal salary for improved student outcomes. This is a type of reward often determined by a performance based evaluation, for example assessing either individual teacher’s student outcomes or student outcomes achieved by groups of teachers (Harvey-Beavis, 2003). Performance based incentives may create incentives for better performance but could have a selection effect from attracting better qualified teachers to the profession (Hoxby & Leigh, 2005). Accordingly, there are individual and group based incentives based on objective and subjective (manager/employer assessment) evaluation criteria.

Group incentives (financial rewards) for teachers are the main focus of this paper. There is an ongoing debate in literature about the merits of group incentives compared to individual incentives. The major aim of the review is to determine the effect of group incentives on student achievement and effects on other outcomes like teacher behavior. Consequently, we examine the current theoretical arguments and empirical evidence on the effectiveness of group incentives for teachers. This paper adds value to literature by including the latest studies from both OECD and developing countries. The paper also employs a systematic review approach in documenting evidence from recent empirical studies in teaching. We also complement our findings with key insights from the private sector literature. The findings of the review are intended to inform policymakers on the current evidence, policy implications of such evidence and identify areas requiring further research.

Theoretical predictions on whether group incentives are more effective than individual incentives are ambiguous. Group incentives can either cause free riding and negatively impact performance or lead to similar or greater results as individual rewards under certain conditions like smaller group sizes where peer monitoring and
complementarities are enhanced. We find that empirical evidence on the effectiveness of group incentives in teaching is limited and that this is a nascent but growing field in literature. The empirical review found six studies on teaching and eight studies on private firms. There is mixed empirical evidence on the effect of group incentives with some studies finding no effect and others finding significant and substantial positive effects on student outcomes. There are differences between studies from developing countries which find positive effects and studies from developed country studies which find mixed effects implying possibly different policy directions. However, we could not draw robust conclusions, as it was not possible to carry out a meta-analysis of empirical evidence given the diversity of student outcomes, design and methodological approaches used in the studies. Evidence from private firms points to positive effects of group incentives on productivity. However we also find that private sector studies are less robust in design compared to the studies on teaching.

The paper is organized as follows. Section II discusses theoretical arguments that can be used to explain the effects of group incentives in teaching. Section III describes the methods used to search the relevant literature and compile the review. Section IV and V describe the empirical studies and the results, while section VI presents the policy conclusions of the study.

Theoretical Framework

In this section we identify theories that can be used to predict the effects of group incentives. We begin by describing theories used to explain the effectiveness of a general incentive and these include the principal agent theory and the theory of intrinsic motivation. We then summarize the arguments found in literature concerning the effectiveness of group incentives compared to individual incentives.
In economic theory, the principal-agent theory embodies the economic rationale for incentives (Umansky, 2005; Fama & Jensen, 1983). In the principal agent theory the school (principal) wants teachers (agents) to work for the school’s best interests. However, incentive conflict occurs when their goals are not aligned (Froeb & McCann, 2009). With incentive conflict there are two problems which arise from asymmetric information: adverse selection (affecting hiring decisions) and moral hazard (affecting teacher motivation and performance) (Froeb & McCann, 2009). One way to solve the moral hazard problem is through external or extrinsic motivation like incentives. For instance incentives like bonuses can help unite the goals of the school and the teacher, where the teacher is motivated to increase work effort and boost performance. Still incentives may fail to address the moral hazard problem especially in a multi-tasking environment like teaching where an incentive could lead to negative effects on teacher behaviour. While teachers engage in multiple tasks such as teaching the curriculum, classroom monitoring and control and preparing students test, teacher incentives are often linked to student test scores at the expense of other student outcomes. Hence, the multi-task moral hazard problem would lead to reduced effort and sub-optimal outcomes in the other tasks (Muralidharan & Sundararaman, 2011). Furthermore weak measurements and evaluations make incentives susceptible to manipulation and ‘gaming’ (Holmström & Milgrom, 1991; Prendergast 1999). Some of the examples of sub-optimal teaching behavioural responses to incentives found in literature include ‘teaching to the test’ (Holmstrom & Milgrom, 1991), excluding weaker students from testing (Jacob, 2005), short term tricks to boost test scores and cheating as experiences in USA with the No Child Left Behind Act show (Figlio & Winicki, 2005; Jacob & Levitt, 2003).

The principal-agent theory has also been criticized for failing to address intrinsic motivation and within this theory incentives can undermine worker effort (Holmström & Milgrom, 1991; Kohn, 1993). Intrinsic motivation is a theoretical mechanism rooted in
psychological theory that can explain teacher behaviour and student outcomes in response to general incentives. According to Deci (1971) one is intrinsically motivated if they perform a task with no expectation of reward except enjoyment in the task itself. Intrinsic motivation is viewed as a pathway to achieving greater performance. In the field of teaching it is especially crucial as it is suspected that teachers choose the profession because of high intrinsic motivation which may gradually dissipate over time if teachers feel underappreciated for their efforts (Muralidharan & Sundararaman, 2011; Mullainathan, 2006). Hence incentives are often given as a source of extrinsic motivation. However, the theoretical predictions of how incentives could influence intrinsic motivation of teachers and thus affect student outcomes, is ambiguous. Financial incentives are mostly seen as extrinsic motivation that may crowd out intrinsic motivation thereby affecting the intended outcomes (Deci, Koestner, & Ryan 1999; Fehr & Falk, 2002). Two psychological processes-self determination and self esteem-determine the psychological conditions under which a crowding out or crowding in effect of incentives occur (Frey & Jegen, 2001). For instance, when individuals feel that an incentive undermines their self determination and self esteem, they reduce their intrinsic motivation in what is referred to as a crowding out effect. The opposite response, a crowding in effect occurs if the incentive is perceived as supportive and reinforcing of one’s norms and self esteem is strengthened and self determination improved (Frey & Jegen, 2001). Therefore how an incentive is designed and communicated is important as the way an incentive is perceived can influence its impact (Mularidharan & Sundararaman, 2011).

The basic prediction of the principal-agent theory is that an incentive would directly increase performance through a price effect, achieved by either increasing the marginal monetary benefit of working harder or by increasing the marginal cost on moral hazard (Frey & Jegen, 2001). When the principal agent theory and intrinsic motivation theory are
considered together, there is likely an interaction between the crowding out effect and the price effect of a monetary incentive (Frey & Jegen, 2001). Gneezy & Rustichini (2000) use an experiment to demonstrate whether behavioural response to a monetary incentive is dependent on the crowding out effect of intrinsic motivation dominating the price effect of the incentive. They find that whenever the crowding out effect dominates the price effect, work effort decreases and only starts to increase once intrinsic motivation has been crowded out completely (Frey & Jegen, 2001; Gneezy & Rustichini, 2000).

As the theories reviewed so far have shown, both psychological and economic theories justify incentives on the premise that they provide motivation for increased performance. Yet, in the private sector and in education, group incentives are more prevalent than individual incentives (Lavy, 2002). In the private sector this is explained by the limitations of individual incentives that firms try to avoid. Individual incentives often impose extra risks on employees which ultimately incur higher wage costs on the firm (Lavy, 2002; Prendergast, 1999). The occurrence of partial contracts, may lead to sub-optimal behavioural responses from workers focusing only on the aspects of performance that are rewarded (Prendergast, 1999). In education, the prevalence of group incentives is attributed to the inherent teamwork in education and the difficulty in measuring individual contributions (Lavy, 2002). This is validated by an important insight from the principal-agent theory, that group incentive pay is more likely used when co-ordination across employees increases, there is imperfect observation of individual output but difficult observation of individual effort (Larkin, Pierce & Gino, 2011). This is unlike individual incentive pay which is more likely to be used when individual output is observed better than actual effort or when employee skills are unobservable and there is a growing need for sorting workers by skill (Larkin et al., 2011).
Still, there is an ongoing debate in literature about whether group incentives are more effective than individual incentives. On the one hand, the basic argument is that group incentives promote free riding, especially in large groups where it’s difficult to monitor effort and therefore yield inferior outcomes to individual incentives (Mularidharan & Sundararaman, 2011; Milanowski, 2007; Holmstrom & Milgrom, 1991). On the other hand, group incentives can yield similar outcomes as individual incentives especially if teachers are able to easily monitor each other like in small schools. Group incentives can lead to better outcomes than individual incentives in cases where there are clear gains to collaboration and complementing each other in output (Azordegan et al., 2005). Larkin et al (2011) premise their theoretical predictions about the efficacy of group incentives vs. individual incentives on an integrated framework of the agency theory and psychological costs of compensation, specifically highlighting two important psychological factors; social comparison processes and overconfidence. Social comparison theory argues that individuals self-evaluate their capability and opinions by comparing with their peers (Festinger, 1954; Larkin et al., 2011). Larkin et al (2011) predict that the costs of social comparison would reduce the effectiveness of individual incentive pay leaving the firm to either impose wage secrecy which is ineffective or reduce the use of performance-based incentives. However, a frequent solution in such situations is the implementation of group incentive pay. Group incentives would then reduce the costs of social comparison especially when individual effort is fairly homogenous within a team. Overconfidence refers to individuals expressing a high degree of optimism in their predictions, overestimating ability and ability relative to others (Vallone, Griffin, Lin & Ross, 1990; Larkin et al., 2011). Overconfident workers usually self select into individual incentive pay self select into inaccurate performance based positions that do not match their skill set. As a result, workers who overestimate their speed of productivity tend to expect greater pay than they will receive which may lead to a high turnover within the firm as workers leave.
their jobs to seek greater compensation elsewhere (Buehler, Griffin & Ross, 1994; Larkin, Pierce & Gino, 2011; Larkin & Leider, 2011). Larkin et al (2011) predict that in settings where there is strong social comparison, group incentives are strategically used to reduce the costs of overconfidence and this would only be effective if the individual contribution of group members is observable.

The theoretical review has increasingly highlighted how theory is ambiguous on the effects of incentives in general and additionally on whether group incentives are more effective than individual incentives. There are several key takeaway points from theory. First, incentives generally affect intrinsic motivation thereby leading to negative or positive performance outcomes. Second, incentives have a price effect which tends to lead to positive outcomes, but this may be dominated by the crowding out effect on intrinsic motivation. Third, group incentives may be more preferable to individual incentives in reducing the costs of overconfidence and social comparison. Finally, group incentives can either cause free riding which negatively impacts performance or under certain conditions (peer monitoring and complementarities) they can lead to similar or greater performance as individual rewards.

**Method**

*Defining Group Incentives*

In the private sector there are various forms and types of group rewards or incentives utilized. They include bonuses, profit sharing, gain sharing, tournament or competition based schemes, piece rate incentives schemes, and target based schemes offering both incentives and penalties (Nalbantian & Schotter, 1997). We restrict our review to empirical studies focusing on group performance linked salary bonuses e.g. tournament incentives, school-wide incentives and team based incentives. Tournaments rank schools based on student achievement and hence the award is given based on performance relative to others (Ladd, 1999; Lavy, 2002). Tournament incentives have the advantage of being cost-effective as
prizes are fixed based on relative performance. They also are beneficial as relative performance can be used as a standard in the absence of clear standards that can be used (Lavy, 2002). School wide incentives refer to rewards linked to school average test scores rather than individual teacher test scores. Team based incentives refer to rewards linked to team average test scores within a school irrespective of the school average test scores e.g. grade five teachers within a school.

Search Strategy

We searched for relevant published empirical studies and working papers from May 2011 to September 2011. The following electronic databases and search engines were searched: Science Direct, JSTOR, Social Science Research Network, Economic Papers, Wiley, Google Scholar and the Google search engine. The following keywords were used: “teacher incentives”, “collective incentives”, “group incentives teaching”, “teacher bonus”, “performance related pay”, “group incentives”, “incentive pay”, “team incentives”, “incentives and workers” and “performance reward”. We searched first for articles on teaching and then for articles on firms. Our search found 15 articles on teaching including two earlier literature reviews (Podgursky & Springer, 2007; Harvey-Beavis, 2003). As for literature on firms, we initially found 15 studies. Studies had to meet the following inclusion criteria: 1) Studies had group incentives as the primary independent variable. 2) They were randomized controlled trials, quasi-experimental studies or non-experimental studies (with inferential statistics). 3) Study assessed student achievement outcomes (education) or productivity related outcomes (firms). 4) They compared outcomes of treated and control groups or at least utilize a before and after comparison (of a single group). 5) Studies used credible inferential statistical analysis for evaluation.

Since our primary focus is on teaching, and most of the studies on teaching were from the late 1990s and 2000s, we excluded studies from before 1990. We also excluded studies
that mainly assessed individual incentives and relied on descriptive statistics. After applying our inclusion criteria, we identified five studies on teaching and eight studies on firms.

**Outcome Measures**

Student test scores are the main outcome for student achievement for studies on teaching. Secondary outcomes include dropout rates, graduation rates, exam participation, test taking, teacher attendance and teacher turnover. The primary outcome for studies on firms is productivity or output.

**Methodology Assessment**

The gold standard for impact evaluation is the experimental study. Hence we consider randomized or experimental studies to have the least risk of bias, quasi-experimental studies to have moderate risk of bias while non-experimental studies have the highest risk of bias. We also appraised the quality or validity of the control group. Studies with a control group approximating the counterfactual were considered to have least risk of bias. Therefore the methodological appraisal relied on the following criteria; whether i) the study was experimental, quasi-experimental or non-experimental; ii) there was a valid control group. Scores were assigned in assessing risk of bias based on the criteria and they are: one (low risk), two (moderate risk) and three (high risk). Total methodological strength was rated from one (weakest methodology) to three (strongest methodology/design). We used a narrative systematic review since the studies vary in designs, outcome measures and statistical reporting.
Description of Studies

Characteristics of Studies

Table A1 (see appendix) provides details of the country, study population, type of group incentive and outcomes for each study on teaching incentives. All of the six studies looked primarily at student outcomes. Four studies were based at primary schools (Goodman & Turner, 2010; Muralidharan & Sundararaman, 2011; Glewwe, Ilias & Kremer, 2010; Ladd 1999). Fryer (2011) and Lavy (2002) assessed incentives at middle and high schools. Two of the studies were rural based and from developing countries, India and Kenya (Muralidharan & Sundararaman 2011; Glewwe et al., 2010). Three others are from the U.S.A and were urban based (Fryer 2011; Goodman & Turner, 2010; Ladd, 1999). One study was nationally representative and from a middle income country (Lavy 2002). Five studies assessed school wide monetary incentives (Fryer, 2011; Goodman & Turner, 2010; Muralidharan & Sundararaman, 2011; Lavy, 2002; Ladd, 1999). One study assessed a school wide in-kind incentive (Glewwe et al., 2010). The amounts of the rewards ranged from as low US$200-715 (Lavy, 2002) to as high as $3000, a 4.1% of average annual salary in Fryer (2011) and Goodman & Turner (2010). In Fryer (2011) and Goodman & Turner (2010), schools that met only 75% of the target performances received a reward of $1500 per teacher. Lavy (2002) and Ladd (1999) assessed tournament based incentives where a rank-order approach is followed in performance with incentive awarded based on rank. In the study by Lavy (2002) the cash incentives were distributed to the top 33% of schools according to a ranking in terms of relative performance, with the highest scoring school winning $105,000 and the lowest winning $13,250 of which 75% of the award was given to teachers as a salary bonus. In the study by Ladd (1999) the top 20% of schools received the award of US$ 1000 per teacher while the next 30% were given $425 US per teacher.
One study compared the effects of group incentives to individual incentives and an alternate program of increasing resources (Muralidharan & Sundararaman, 2011). Lavy (2002) compared the effects of group incentives to an alternate program of increasing resources. Student test scores are the common outcome for the five studies, and are primary performance measure linked to the incentives. Other secondary outcomes assessed include dropout rates (Glewwe et al., 2010; Ladd, 1999), exam participation (Glewwe et al., 2010; Lavy, 2002), graduation rates (Lavy, 2002; Fryer, 2011). All studies with the exception of Lavy (2002) assessed teacher attendance or presence.

Table A2 in the appendix describes the country, study population, type of group incentive and outcomes for studies on private firms. The studies are based on incentives implemented in small scale firms, corporations and experimental games. Among the eight selected studies, three types of incentives are assessed. Two studies evaluated firm wide group incentives (Dixit & Paul, 2010; Knez & Simester, 2001). Four studies evaluated unit or group based incentives within a firm (Boning, Ichniowski & Shaw, 2007; Bhattacherjee, 2005; Hamilton, Nickerson & Owan, 2003; Hansen, 1997). Two studies conducted experiments comparing tournament or competition based incentives to group and individual incentives (Nalbantian & Schotter, 1992; Erev, Bornstein & Galili, 1993). Unlike the teacher incentives which were based on annual performance, the incentives in the firms were based on monthly performance. Productivity is the primary outcome in all of the studies with the exception of Dixit & Paul (2010) who use return on investment.

Methodological Quality of the Studies

Table 1 presents the empirical summary of the studies on teachers’ group incentives. Of the six studies, four utilized field experiments with valid control groups and are rated high on methodological strength (Muralidharan & Sundararaman, 2011; Goodman & Turner, 2010; Fryer, 2011; Glewwe et al., 2010). One study utilized a quasi-experimental design, but
the control group was not based on a proper counterfactual and was assigned a moderate score on methodological strength (Lavy, 2002). One study was non-experimental (Ladd, 1999). The sample size among the five studies ranges from 84 to 1122 schools. All six studies followed a panel design. All studies used multivariate regression methods in statistical analysis.

Of the eight studies from the private sector, two utilized experimental games with students (Nalbantian & Schotter, 1992; Erev et al., 1993). However both studies did not have the ideal control group based on the counterfactual of no incentive and hence the studies were assigned a moderate score on methodological strength. The remaining six studies were non-experimental. Three of these had valid control groups (Dixit & Pal, 2010; Boning et al., 2007; Knez & Simester, 2001). The other three did not have a proper counterfactual group and hence score the least on methodological strength (Bhattacherjee, 2005; Hamilton et al., 2003; Hansen, 1997). All of the studies with exception of the two followed a time series design. The experimental game theory studies followed a cross sectional approach (Nalbantian & Schotter, 1992; Erev et al., 1993). All the studies employed multivariate regression methods in statistical analysis. Overall, studies on teaching utilize robust designs (experimental and quasi-experimental) than studies from the private sector.

**Findings**

*Group Incentives for Teachers*

As seen in table 1, the four experimental studies had mixed results on the effect of group incentives on student test scores. The two experimental studies from India and Kenya found a significant and positive impact of group incentives on student scores in the range of 0.14 standard deviations (SD) to 0.16SD (Glewwe et al., 2010, Muralidharan & Sundararaman, 2011). The other two experimental studies by Fryer (2011) and Goodman and Turner (2010) did not find significant effects on test scores. The quasi-experimental
study by Lavy (2002) found a positive and moderate effect on test scores in the first year which rose significantly in the second year to 1.75 points (secular schools) and 3 points (religious schools). The non-experimental study (Ladd, 1999) found a positive effect of group incentives on math test scores of up to 17% and reading test scores of up to 15%. However, Ladd (1999) also found racial disparities in the test scores with improved test scores for Hispanic and White students and a decline for Black students. In comparing group incentives and individual incentives, Muralidharan & Sundararaman (2011) found that individual incentives had larger effects than group incentives. However both types of incentives performed better than an alternate program of increasing school resources. Lavy (2002) found that group incentives are more cost effective compared to increasing school resources.

Studies also looked at other outcomes in learning, teacher behaviour and quality. Other effects of group incentives include an increase in test scores of non-incentive linked courses and teacher effort (Muralidharan & Sundararaman, 2011), improved exam participation and preparation (Muralidharan & Sundararaman, 2011; Glewwe et al., 2010; Lavy, 2002), increased graduation rates (Lavy, 2002), improved test taking skills (Glewwe et al., 2010). However Glewwe et al (2010) did not find any significant effects on the test scores of non-incentive courses, dropout rates, teacher attendance, homework assignment or pedagogy. Muralidharan & Sundararaman (2011) did not find any effect on teacher attendance, homework assignment and pedagogy. Goodman & Turner (2010) accounted for free riding in their analyses and find that group incentives le to free-riding which reduced the influence of the incentives. They also found that a small reduction in teacher absence for the largest incentives of $3000, especially in schools with fewer teachers. However they did not find any effect on teacher turnover or qualifications of newly hired teachers. Only one study assessed the post-program effects of group incentives (Glewwe et al., 2010) and found no sustained gains after the end of the group incentives.
**Group Incentives in Firms**

All the non-experimental studies found a significant effect of group incentives on productivity ranging from 0.493% (Boning et al., 2007) to 18% (Hamilton et al., 2003). Dixit and Pal (2010) found an increase of 0.03% in return on investment (see table A2 in appendix). The two experimental studies (Nalbantian & Schotter, 1997; Erev et al., 1993) compared tournament incentives to individual and group incentives. Nalbantian & Schotter (1997) found that tournament incentives performed better than group incentives in the earlier stage of the game, however by the later stages of the game there was no significant difference as a shirking equilibrium was reached. While individual incentives performed better than group incentives they were the least cost effective. Erev et al (1993) found that individual incentives had larger effects on productivity than tournament incentives. Tournament incentives in turn had larger effects than group incentives. Over the duration of the experimental games, Erev et al (1993) found an increase in productivity from tournament incentives, a decline in productivity from group incentives and no change in productivity from individual incentives.

**Discussion**

**Policy Implications**

The majority of the empirical studies in teaching show overall short term gains in test scores. However, positive impacts are mostly reported in developing countries and middle income countries compared to developed countries. In the USA, one non-experimental study finds positive effects of group incentives and two experimental studies show no effect. There are several possible reasons why some studies, especially from the USA, found no significant impact of group incentives on student test scores compared to studies conducted in developing countries. First, the incentive designs in the USA are likely more complex partly due to union influence and hence provided teachers with less agency compared to developing
countries (Fryer, 2011). Second, as predicted by theory, the costs of free riding could be outweighing the costs of co-ordination or teamwork. In teaching, there is evidence that free riding could be minimizing the effect of group incentives especially in schools with a large number of teachers. Goodman & Turner (2010) found a small decline in teacher absence in schools with fewer teachers; however this positive change in teacher effort was not large enough to increase student test scores. Empirical studies on private firms also confirm the role of free riding as they show that shirking increases over time under group incentives (Nalbantian & Schotter, 1992) and a small group size could respond more positively to a group incentive than a large one (Bhattacherjee, 2005). Third, the size of the incentives could also be a factor. As mentioned earlier, the three empirical studies from the USA all evaluated similar valued incentives and reported different effects. Goodman & Turner (2010) allude to the possibility of larger incentives having a greater effect than small incentives as evidenced by the small decline in teacher absence in schools with the largest incentives of $3000 compared to $1500. Yet, increasing the size of incentives might be problematic as it may increase the risk borne by teachers resulting in the cheating or gaming of the system (Glewwe et al., 2010). Interestingly, Muralidharan & Sundararaman (2011)’s study in India could not determine whether the differences in test scores disaggregated by teacher base salary could be attributed to the size of the incentive or teacher experience. Their study finds that higher base pay respond less positively to group incentive pay compared to teachers with lower base pay as the incentive would be a larger share of lower paid teachers. However, the experienced teachers tend have higher base pay compared to younger teachers who have been shown to respond better to policy initiatives than experienced teachers. Finally, the ineffectiveness of group incentives regarding student test scores in the USA could also be attributed to several other factors yet to be explored empirically in the education sector. One factor is the crowding out of intrinsic motivation where it is possible that teachers feel group incentives
are controlling and undermine their self determination and self esteem. Another important factor is team composition and past experience of the team. In private firms, teams receiving group incentives and composed of high ability workers have greater productivity than others (Hamilton et al., 2003). Nalbantian & Schotter (1997) showed that past positive group experience is correlated with increased productivity. Hence, from a policy perspective, it appears that group incentives may have different effects in different contexts, and that incentive design, incentive size and free riding may matter more in some contexts than others. The mixed results from the studies conducted in the USA, make it difficult to recommend policy directions for developed and middle income countries.

There is mixed empirical evidence on the impact of group incentives on other learning and educational outcomes. While most of the studies from developing countries report positive outcomes in exam participation and preparation, test taking skills and teacher effort, there are mixed results on the potential spillover effects to non-incentive linked courses (increase in Muralidharan & Sundararaman, 2011 but no effect in Glewwe et al., 2010) and mixed results on dropout rates (Glewwe et al., 2010 found no effect; Lavy, 2002 found a decrease). The studies also found no incentive effect on teacher attendance, teacher turnover, homework assignment or pedagogy. In the USA however, studies find a small effect on teacher attendance but little effects on graduation rates, turnover and qualifications. It would appear that group incentives generally have a positive effect on other learning and teaching outcomes. This raises the question whether group incentives should only be linked to test scores as a performance measure? Again, it would seem context matters. In developing countries test scores may still work as there are critically low levels of learning and any increase in classroom teaching effort or improvement test taking skills might lead to better broader learning outcomes. In developed countries however, where empirical studies are still largely conducted in the USA the jury is still out on this matter.
An important policy consideration in choosing group incentives has to do with selecting an effective design. In the private sector literature, Nalbantian & Schotter (1997) and Erev et al (1993) showed that tournaments are more effective than group incentives. In this paper we review empirical studies focusing on two types of group incentives for teachers e.g. tournament-based incentives and school-wide incentives. Two studies which evaluated tournament-based incentives found positive learning outcomes, but these studies were non-experimental (Lavy, 2002; Ladd, 1999). An inspection of the other non-tournament based studies shows that some reported positive test scores (Muralidharan & Sundararaman, 2011; Glewwe et al., 2010) while the two experimental studies from the USA did not find significant effects. Further empirical evidence is required on this matter, especially to determine if tournaments may be the most effective design of teachers’ group incentives in a developed country like the USA. The financial nature of incentives is another element for consideration. In this paper we review only one study that evaluated in-kind incentives; hence there is limited evidence on whether in-kind incentives would be more effective than monetary incentives, which might be an area of interest to education policy in developing countries.

Our empirical review continues the debate on whether group incentives are preferable to individual incentives in education. Muralidharan & Sundararaman (2011) found that individual incentives significantly outperform group incentives, yet this is the only study available in the education field. Experimental games in the private sector literature show that individual incentives outperform group incentives, but are less cost effective (Nalbantian & Schotter 1997, Erev et al., 1993). With this limited evidence, we are unable to establish firm conclusions on the matter. A positive sign however is that group incentives perform better
and are more cost effective than input based incentives like increasing school resources (Muralidharan & Sundararaman, 2011; Lavy, 2002).

Most of the evidence discussed so far focuses on short run gains in learning or lack thereof. Of equal importance is whether group incentives lead to long term gains in student achievement. One study shows no long term gains after end of incentive program in Kenya (Glewwe et al., 2010). Could this be evidence of ‘gaming’ or manipulation by teachers for short term gain? The answer to this question is unclear as the program in question was designed to discourage manipulation and gaming. The schools were discouraged from forcing weaker students to repeat, drop out, or not take the exam and students who did not take the exam were assigned low or zero test scores. The study’s test score calculation only included students enrolled in school at the start of term as a way of deterring schools from recruiting strong students. While there could be no long term gains from group incentives, short term gains in test scores may still be policy relevant. Improved test scores for those graduating could translate into positive returns on long term human capital accumulation and labor market returns (Muralidharan & Sundararaman, 2011). Still, further empirical evidence is required on this issue.

Research Implications

We could not carry out a meta analysis due to the diverse incentive designs, durations of implementation and outcomes assessed. Our review concludes that there is a nascent evidence base on teaching with a growing number of emerging experiments unlike in private firms were studies are based on less rigorous non-experimental designs. However there is still a need for further experimental evidence in the education sector in order to draw firm conclusions on the effectiveness of group incentives. Future experimental studies could also capture data on teacher attitudes and motivation which may shed light on aspects like intrinsic motivation, social comparison and overconfidence. We also recommend that future
experimental designs link group incentive pay to non-test score outcomes like teacher behavior and quality.

**Conclusion**

Current empirical evidence shows that group incentives for teachers have significant positive effects on student test scores mostly in developing countries compared to developed countries, hence context may be crucial in determining policy direction (Dixit, 2002). Concerning developed countries, the positive impacts on student outcomes are mostly found in non-experimental studies making it difficult to recommend policy directions for developed and middle income countries. While there is limited evidence on the long term gains from group incentives, short term gains in test scores especially for graduating students may have long term positive labor market returns. Current evidence has unclear policy implications on the long term learning gains from group incentives. It is also difficult to ascertain whether the key to positive effects is the size of the incentive, group size, teacher intrinsic motivation, type of incentive (rank type vs. non-rank type) or design (test score based incentives). Therefore there is a vital need for further experimental studies on the effectiveness of teachers’ group incentives to help establish firm conclusions and policy recommendations.
References


Larkin, I., Pierce, L., & Gino, F. The Psychological Costs of Pay-for-Performance:


TABLE 1
Empirical summary of the effects of group incentives on student achievement and other outcomes

<table>
<thead>
<tr>
<th>Study</th>
<th>Country</th>
<th>Study Design</th>
<th>Test scores</th>
<th>Other outcomes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Goodman &amp; Turner (2010)</td>
<td>New York, USA</td>
<td>Experimental</td>
<td>No effect</td>
<td>Group bonuses led to free-riding</td>
</tr>
<tr>
<td></td>
<td>Urban Elementary</td>
<td>School wide</td>
<td>$3000</td>
<td>Schools with &lt;5 teachers in tested classrooms, ↑teacher attendance by 0.5days/teacher leading to 2.5 less absences over 5 months</td>
</tr>
<tr>
<td></td>
<td>schools</td>
<td></td>
<td>$1500</td>
<td>No effect on teacher turnover and qualifications</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(for 75%</td>
<td>↑ in test scores of non-incentive courses</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>performance)</td>
<td></td>
</tr>
<tr>
<td>Muralidharan &amp;</td>
<td>India Rural</td>
<td>Experimental</td>
<td>↑0.16SD (</td>
<td>Incentives more cost effective than increasing school resources</td>
</tr>
<tr>
<td>Sundararaman (2011)</td>
<td>Primary schools</td>
<td>School wide</td>
<td>group)</td>
<td>No effect.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>↑0.27SD (</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>individual)</td>
<td></td>
</tr>
<tr>
<td>Fryer (2011)</td>
<td>New York, USA</td>
<td>Experimental</td>
<td>No effect</td>
<td>No effect.</td>
</tr>
<tr>
<td></td>
<td>Urban Elementary</td>
<td>School wide</td>
<td>$3000 (&lt;4.1% of average annual salary)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>middle and high</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>schools</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Study</td>
<td>Country</td>
<td>Study Design and Type of Incentive</td>
<td>Test scores</td>
<td>Other outcomes</td>
</tr>
<tr>
<td>---------------</td>
<td>------------------</td>
<td>-----------------------------------</td>
<td>-------------</td>
<td>--------------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Glewe et al.</td>
<td>Kenya Rural Primary schools</td>
<td>Experimental School-wide In-kind 21-43% of monthly salary</td>
<td>0.14SD</td>
<td>↑ exam participation ↑ test taking skills ↑ test preparation sessions. No effect on non-incentive linked courses, exam participation, dropout rates, teacher attendance, homework assignment and pedagogy.</td>
</tr>
<tr>
<td>Lavy (2002)</td>
<td>Israel High schools</td>
<td>Quasi-experimental School-wide (tournament) $200-$715</td>
<td>1.75 points (secular schools) 3 points (religious schools)</td>
<td>Increase in number of credit units ↑ graduation rates ↑ exam participation. Group incentives more cost effective than increasing resources ↑ Hispanic and white test scores ↓ Black students test scores ↑ principal turnover</td>
</tr>
<tr>
<td>Ladd (1999)</td>
<td>Dallas, Texas Urban grade 7 schools</td>
<td>Non-experimental School-wide (tournament) $1000 (top 20%) $425 (next 30% of schools)</td>
<td>17% (math) 15% (reading)</td>
<td></td>
</tr>
</tbody>
</table>

Notes: SD refers to standard deviation, < refers to less than, ↑ refers to increase, ↓ refers to decrease.
### Table A1

**Studies included in the review: Group incentives in teaching**

<table>
<thead>
<tr>
<th>Study</th>
<th>Country and Study Population</th>
<th>Study Design and Time Span</th>
<th>Type and Amount of Group Incentive</th>
<th>Outcome variable</th>
<th>Findings</th>
</tr>
</thead>
<tbody>
<tr>
<td>Muralidharan &amp; Sundararaman (2011)</td>
<td>India Rural Primary schools, Grades 2-5 Sample size- 200 treated (100 team and 100 individual incentives), 300 control (200 with increased resources)</td>
<td>Panel Quantitative analysis- Fixed effects regression 2005-2006 Control group: 1 Experimental: 1 Methodological strength: 3</td>
<td>School wide bonus Individual bonus 3% of annual salary</td>
<td>Test scores, teacher attendance, teacher effort</td>
<td>Both incentives ↑math scores (0.28SD), language (0.16SD), combined subjects (0.22SD) Group incentives- math (0.23SD), language (0.09SD), combined subjects (0.16SD) Individual incentives-math (0.32SD), language (0.22SD), combined subjects (0.27SD) ↑ in teacher effort and non-incentivized courses. No effect on teacher attendance, pedagogy. Incentives more cost effective than increasing school resources</td>
</tr>
<tr>
<td>Study</td>
<td>Country and Study Population</td>
<td>Study design and Time Span</td>
<td>Type and Amount of Group Incentive</td>
<td>Outcome variable</td>
<td>Findings</td>
</tr>
<tr>
<td>---------------</td>
<td>------------------------------</td>
<td>----------------------------------------------------------------------------------------</td>
<td>-----------------------------------</td>
<td>----------------------------------------------------------------------------------</td>
<td>-------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Fryer (2011)</td>
<td>New York, USA Urban Elementary, Middle and High schools, Sample size- 233 treated and 163 control schools</td>
<td>Panel Quantitative analysis- Regression, IV 2007-2010 Control group:1 Experimental: 1 Methodological strength: 3</td>
<td>School-wide $3000 (&lt;4.1% of average annual salary)</td>
<td>Test scores, attendance, graduation rates, student and teacher behavior</td>
<td>No significant effect on any outcomes.</td>
</tr>
<tr>
<td>Glewe et al.  (2010)</td>
<td>Kenya Rural Primary schools, grades 4-8 Sample size- 50 treated and 50 control schools</td>
<td>Experimental Panel Quantitative analysis-linear probability model, random effects probit regression 1997-2000 Control group:1 Experimental: 1 Methodological strength: 3</td>
<td>School-wide In-kind incentives Valued at 21-43% of monthly salary</td>
<td>Test scores, exam participation, dropout rates and teacher response</td>
<td>↑ in test scores in year 2 by 0.14SD. Significant ↑ in participation in incentive linked exams, test taking skills, test preparation sessions. No significant effect on non-incentive courses, exam participation, dropout rates, teacher attendance, teacher presence, homework assignment and pedagogy. No post program effect</td>
</tr>
<tr>
<td>Study</td>
<td>Country and Study Population</td>
<td>Study design and Time Span</td>
<td>Type and Amount of Group Incentive</td>
<td>Outcome variable</td>
<td>Findings</td>
</tr>
<tr>
<td>---------------</td>
<td>-----------------------------</td>
<td>---------------------------</td>
<td>-----------------------------------</td>
<td>------------------</td>
<td>--------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Lavy (2002)</td>
<td>Israel</td>
<td>Panel</td>
<td>School-wide (tournament, top 33%)</td>
<td>Test scores, course taking, exam participation, dropout rates, graduation rates</td>
<td>Modest effect in first year, Larger ↑ in second year. Test scores (1.75 points in secular schools and 3 points in religious schools). ↑ in number of credit units, graduation rates, exam participation. Comparison program also has positive impact, but group incentives more cost effective.</td>
</tr>
<tr>
<td></td>
<td>High schools (religious and secular schools)</td>
<td>Quantitative analysis - regression discontinuity design, difference in differences, matching 1993-97</td>
<td>$200-$715</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Sample size- 62 treated schools and 22 in alternate program of increasing resources</td>
<td>Control group: 2 (no actual counterfactual of without reward) Quasi-experimental: 2 Methodological strength: 2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Ladd (1999)</td>
<td>Panel</td>
<td>School-wide (tournament, top 20%)</td>
<td>Math test scores, Math and reading test scores, dropout rates, principal turnover</td>
<td>7th grade math test scores ↑ from 14% to 117%, reading (from 10% to 15%). Hispanic and white students ↑ test scores, but ↓ for black students. Principal turnover ↑</td>
</tr>
<tr>
<td></td>
<td>Dallas, Texas grade 7 schools</td>
<td>Quantitative analysis-difference estimation 1991-1995</td>
<td>$1000</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Treated-Dallas schools</td>
<td>Control group: 1 Non-experimental: 3 Methodological strength: 2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Control-El Paso, Fort Worth, San Antonio and Houston</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Notes: SD refers to standard deviation, < refers to less than, ↑ refers to increase, ↓ refers to decrease.
### TABLE A2

**Studies supplementing the review: group incentives in private firms**

<table>
<thead>
<tr>
<th>Study</th>
<th>Country and Study Population</th>
<th>Study Design and Time span</th>
<th>Type of Group Incentive</th>
<th>Outcome variable</th>
<th>Findings</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dixit &amp; Pal 2010</td>
<td>India 10 small scale firms (auto and furniture)</td>
<td>Time series annual data 1996-2005</td>
<td>Firm wide bonus when output exceeds monthly target</td>
<td>Return on investment</td>
<td>0.03% ↑ in return on investment from group incentive</td>
</tr>
<tr>
<td>Boning et al 2007</td>
<td>USA 19 companies 34 Mini mill steel production lines Treated-30 lines (1372 observations) Control- 4 lines (about 985 observations)</td>
<td>Time series (monthly) Quantitative analysis – Fixed effects regression 1994-97</td>
<td>Group incentive pay/bonus per production line</td>
<td>Monthly productivity (tons of steel)</td>
<td>↑ in productivity by 0.493 percentage points Problem solving teams in combination with group incentive pay further ↑ productivity.</td>
</tr>
<tr>
<td>Bhattacherjee 2005</td>
<td>India Multi business corporation 4 plants No control data, so estimate elasticity of productivity to two types of incentive pay</td>
<td>Time series monthly data Quantitative analysis- Fixed effects estimation 1985-1995</td>
<td>Plant level bonus Department level bonus</td>
<td>Productivity or unit output per worker, monthly absence per worker, labor unit costs per unit output</td>
<td>10% ↑ in incentive pay leads to 0.8% ↑ in productivity Department level incentive has positive effects Plant level incentive has negative effects Smaller group size appears to have less free riding and higher</td>
</tr>
<tr>
<td>Study</td>
<td>Country and Study Population</td>
<td>Study Design and Time span</td>
<td>Type of Group Incentive</td>
<td>Outcome variable</td>
<td>Findings</td>
</tr>
<tr>
<td>----------------</td>
<td>------------------------------</td>
<td>---------------------------</td>
<td>-------------------------</td>
<td>------------------</td>
<td>--------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Hamilton et al 2003</td>
<td>USA Urban Garment manufacturing firm</td>
<td>Time series monthly data 1995-1997</td>
<td>Group piece rate scheme (per unit pay based on collective performance of 6-7 workers in a group)</td>
<td>Weekly productivity/output</td>
<td>18% ↑ in average weekly productivity. Where 4% is due to high ability workers who joined teams and 14% is the team effect. Hence team composition is important.</td>
</tr>
<tr>
<td>Hansen 1997</td>
<td>USA American Express</td>
<td>Time series monthly data 1991-1992</td>
<td>Bonus based on unit performance</td>
<td>Productivity (minutes per telephone call)</td>
<td>17% ↓ in productivity from bonus.</td>
</tr>
<tr>
<td>Study</td>
<td>Country and Study Population</td>
<td>Study Design and Time span</td>
<td>Type of Group Incentive</td>
<td>Outcome variable</td>
<td>Findings</td>
</tr>
<tr>
<td>-----------------------------</td>
<td>------------------------------</td>
<td>----------------------------</td>
<td>----------------------------------------------------------------------------------------</td>
<td>------------------</td>
<td>-----------------------------------------------------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Nalbantian &amp; Schotter 1997</td>
<td>USA</td>
<td>Cross sectional</td>
<td>Tournament based reward, profit sharing, gain sharing, forcing contract (target based schemes) and monitoring schemes (individual incentive)</td>
<td>Mean effort level</td>
<td>In first phase of experiments Tournament/competitive rewards ↑ effort more than forcing contract (target based group incentive). Insignificant in second phase. Shirking equilibrium reached towards end of the game. Positive past group experience leads to ↑ current output levels Intra-firm team competition ↑ group effort Monitoring works but is costly.</td>
</tr>
<tr>
<td>Erev et al 1993</td>
<td>Israel</td>
<td>Cross sectional</td>
<td>Three payoffs- personal reward, group incentive and tournament/competition incentive.</td>
<td>Productivity</td>
<td>On average 376 kg oranges picked under personal reward, 280 kg for group incentive and 380kg under tournament/competition. Over time sharp ↑ in productivity under competition, sharp decrease under group reward and no change under personal reward.</td>
</tr>
</tbody>
</table>

Notes: SD refers to standard deviation, < refers to less than, ↑ refers to increase, ↓ refers to decrease