

Bonuses and performance evaluations

Individual bonuses do not always raise performance; it depends on the characteristics of the job

Keywords: bonus, compensation, variable pay, incentives, performance evaluation, evidence-based bonus design

ELEVATOR PITCH

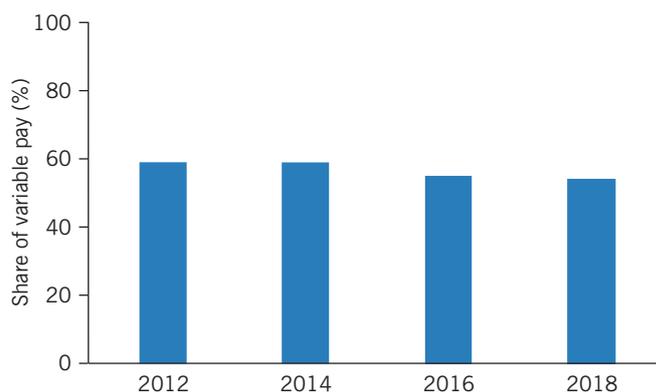
Economists have for a long time argued that performance-based bonuses raise performance. Indeed, many firms use bonuses tied to individual performance to motivate their employees. However, there has been heated debate among human resources professionals recently, and some firms have moved away from individual performance bonuses toward fixed wages only or collective performance incentive schemes such as profit-sharing or team incentives. The appropriate approach depends on each company's unique situation, and managers need to realize that individual bonus plans are not a panacea to motivate employees.

KEY FINDINGS

Pros

- ⊕ Bonuses may raise performance when objective measures are available that assess the key aspects of performance and if workers have sufficient leeway to increase their performance.
- ⊕ Bonuses motivate people who exhibit low task motivation.
- ⊕ Team bonuses can raise performance as peer effects and social preferences can mitigate the so-called free-rider problem.

The prevalence of variable pay has decreased in Germany



Source: Author's own compilation using LPP data. Online at: https://fdz.iab.de/en/Integrated_Establishment_and_Individual_Data/lpp.aspx

IZA
World of Labor

Cons

- ⊖ Bonus schemes can cause distortions when employees work on multiple tasks and efforts for important tasks are hard to assess.
- ⊖ Performance-based bonuses may have limited effects for employees with a conscientious personality.
- ⊖ Bonus payments may affect, and potentially even undermine, other management practices used to motivate employees.
- ⊖ Bonuses based on individual performance often have to rely on subjective performance evaluations and these evaluations tend to be biased.

AUTHOR'S MAIN MESSAGE

Recent empirical evidence on the effects of bonus schemes shows that details of their implementation matter. Bonus schemes may work well for easily measurable tasks when employees have low intrinsic task motivation and performance is hard to monitor. In other contexts, however, there are several factors that limit their benefits. Bonus schemes must thus be carefully evaluated in the specific context of the organization in which they are to be implemented. Firms and academics should pursue evidence-based bonus design by implementing scheme changes for subgroups of employees that allow for robust evaluation of the observed consequences.

MOTIVATION

Economists have traditionally advocated the view that bonuses raise employee performance. The key idea is simple: if people are paid according to their performance, they should be motivated to work harder. Practitioners and academics have sometimes challenged this view. And while for many years the use of performance pay in firms had increased, recent descriptive evidence indicates a potential reversal of this trend (see the illustration on p. 1).

For a surprisingly long time, there was very little clean causal evidence from actual firms on the performance effects of bonuses. A key reason is that if a firm introduces a bonus scheme for all its employees at the same time (as is common in practice) it is virtually impossible to estimate its effects on performance. This is because many other impactful things tend to happen at the same time (business cycle effects, market developments, and so on); any changes in performance can then not be cleanly attributed to changes in the bonus scheme. In a field experiment (randomized controlled trial, RTC), however, a new scheme is only implemented for a subgroup of employees, which allows researchers to cleanly estimate its causal effects. Evidence from a growing number of field experiments and quasi-experiments on the performance effects of bonus schemes in firms is now available and may help shed further light on the subject.

DISCUSSION OF PROS AND CONS

A benchmark case

One of the earliest large-scale quasi-experimental studies investigates the effects of introducing performance pay for technicians repairing windshields at Safelite, the largest US provider of vehicle glass repair services [1]. Safelite technicians had previously worked under fixed hourly wages. Starting in 1993, the firm introduced a performance pay plan which paid technicians according to the number of windshields installed per unit of time. The setting does not correspond to an actual RCT as the new scheme was not randomly assigned to subsidiaries. But it was not implemented at the same time for all and within the rollout period there were subsidiaries working under the old and the new compensation schemes. This created an approximate experimental setting.

The performance pay plan had a very large effect on performance: the number of windshields installed per technician increased by more than 40%. Roughly half of this effect was due to higher performance among legacy Safelite technicians (the “incentive effect”). The other half was a result of Safelite attracting more productive (able) workers under the new performance pay plan (the “selection effect”).

This study constitutes an important benchmark for the literature on performance pay, but it is crucial to think about its specific context. Safelite technicians primarily drove to the customer’s location to conduct windshield repairs. Hence, monitoring employees’ behavior was hard and may have led to a substantial “moral hazard” problem. Moreover, repairing windshields was technicians’ main and essentially only important task. The key metric used to measure performance (the number of windshields repaired) thus measured all key aspects of their importance for the firm. All in all, the Safelite setting was very conducive for performance pay to succeed.

Context factors and design challenges

As shown above, bonus plans may work well in settings where monitoring workers is difficult and where key performance indicators track workers' essential tasks. However, many jobs systematically differ from such settings, which can substantially affect whether and to what extent bonus schemes increase performance. Examples of such alternate settings include:

- In most jobs people work on several tasks and objective performance measures are often not available for all tasks. This may create *multitasking* problems.
- Employees can be motivated by *social preferences* or have *personality traits* that generate intrinsic preferences to do their work well, which may reduce the need to align interests through bonuses.
- Objective performance measures are often only available at the level of a group of employees, potentially giving rise to *free-riding* behavior.
- *Other management practices* such as performance feedback or target setting have also been shown to raise performance; bonus schemes may interact in complex ways with these practices.
- Firms often rely on subjective assessments (for instance by supervisors) to award bonuses which are prone to *biases in subjective performance evaluations*.

Each of these design challenges is discussed below with reference to relevant empirical evidence from field studies and (quasi-)field experiments in firms.

Multitasking

A classical result in the theory of incentives shows that if only subsets of tasks can be measured objectively, and performance pay is based on these tasks, people focus too much on the measurable tasks. A typical example is the question of whether firms should use bonuses to reward produced “quantity” when the “quality” of the output cannot be easily measured. A recent field experiment in several Chinese electronics manufacturing firms finds some evidence in line with this conjecture [2]. In the experiment, workers performing supporting work (such as packaging) in five firms received a sizeable bonus for quantity increases for a few days. Work quality (defect rates) was secretly inspected and showed that the intervention increased quantity but also led to a higher defect rate. Another example where multitasking distortions can play a substantial role is the question of whether bonus payments that reward short-term performance inhibit exploration and creativity. A consequence of such multitasking distortions is that if specific key tasks cannot be easily measured and these tasks are very important to an organization's overall objectives, it may be better not to use performance pay schemes.

Social preferences

Incentive theory in economics has traditionally relied on the *homo economicus* model: the idea that people are guided by their self-interest and hardly care for the well-being of others. However, a large body of evidence in behavioral and experimental economics

(and of course also in psychology) has shown that many people care for the well-being of others and thus have—in the language of economics—social preferences. In the context of bonus plans this has, for instance, been shown in a study with agricultural workers that people tend to take the effect of their actions on their coworkers into account [3]. The study explores a change from a bonus scheme where workers were paid based on their relative performance (they earned more when performing better than their colleagues) to a bonus scheme in which rewards only depended on a worker's own performance. Whereas, in the former, higher efforts by one worker had a negative effect on a coworker's well-being, this was not the case in the latter scheme. The authors find that performance increased substantially after the relative performance component was abandoned. This effect was stronger when coworkers were close friends, which supports the view that social preferences matter. However, the authors provide further evidence that workers “internalize” (i.e. take into account) the effect of their behavior on coworkers only when efforts are mutually observed—which indicates that it is not pure altruism that triggers these effects, but that this social behavior also has a strategic component.

Personality traits

The idea that employees' preferences and traits affect how they respond to monetary incentives has another important implication: individuals who intrinsically feel the obligation to work or those who enjoy the task may respond to a lesser extent to performance pay, as they “try to give their best” even at flat wages.

A recent field experiment among maternity care providers in India supports this view [4]: the authors assessed the personality of the providers measuring the so-called “big five personality traits” (the most widely accepted psychological characterization of personality: openness, conscientiousness, extraversion, agreeableness, and neuroticism). One key personality trait is conscientiousness: people who score high with respect to conscientiousness are those that feel intrinsically driven to do their work or duty well and thoroughly. The study finds that the introduction of performance pay raised workers' performance, but that the positive effect was driven by the less conscientious workers. The highly conscientious workers hardly reacted to the incentive.

Team incentives and free-riding

Objective performance measures are often available only for larger groups of employees. For instance, financial performance is typically assessed only at the level of the whole firm or specific profit center within that firm. Beyond jobs in sales, the profit contribution of individual employees can only rarely be assessed objectively. Performance pay in such settings may thus be based on group outcomes. This can give rise to the so-called free-rider problem, according to which individual employees do not work as hard because they personally receive only a small part of the performance gains from increased effort. Economists have traditionally been very skeptical toward the usefulness of team incentive schemes. However, a growing number of empirical studies support the view that free-riding may be less severe than previously thought. One key reason is again the prevalence of social preferences, which lead employees to internalize the (positive) effect of their own efforts on colleagues' well-being when team bonuses are in place. Another core

mechanism is peer pressure: under team compensation, team members have an incentive to motivate or punish free-riding colleagues.

One study explores the performance effects of a very simple bonus scheme Continental Airlines introduced to improve on-time departure rates [5]. The scheme paid a specific amount to all Continental employees if the airline ranked high in on-time performance among US airlines. Theoretically, this scheme should have led to a huge free-rider problem, as the impact of an individual employee at a specific airport on the entire airline's on-time ranking would be very small. However, the authors provide compelling evidence that the scheme worked by comparing performance changes around the introduction of the scheme between airports in which Continental's operations were performed by its own employees to those where operations were outsourced to separate companies, showing that the performance gains were driven by the former.

Researchers studying the introduction of teamwork in a US garment factory (which coincided with a shift from individual to team incentives) find that teamwork was associated with an increase in productivity [6]. Workers in this setting could initially volunteer to join teams. Interestingly, in particular, high-ability workers decided to join teams early on—even though this often led to wage losses (as these workers had high earnings under individual performance pay). This evidence again supports the view that free-riding may be less prevalent—as low-ability workers had the strongest material incentive to join a team and free-ride. Moreover, it is in line with the view that teamwork can bring about non-monetary benefits, for instance when social preferences outweigh free-rider effects.

A recent field experiment in a large German bakery chain randomly assigned a team bonus to stores and found that the team bonus increased profits significantly [7]. The study also analyzed detailed personnel data to investigate the underlying cause; the evidence shows that the team bonus increased customer serving speed, thereby raising the number of customers served.

Feedback or rewards?

When a firm introduces bonus payments, it not only affects employees through monetary incentives, but it also draws attention to the key metrics used to assess performance. By the same token, the introduction of a bonus may also yield performance feedback that in itself can affect motivation. Evidence in line with this idea comes from a study conducted in a Dutch retailer (selling clothing, shoes, and sports apparel) [8]. The study introduced a team bonus via a six week tournament between different stores. Stores were allocated into groups of five and competed for a fixed bonus paid to all employees in the best performing store in each group. The study finds that this tournament increased sales performance. But strikingly, the effect was strongest and statistically most robust in a “feedback” treatment where the reward (the tournament prize) was not associated with a monetary bonus but merely symbolic.

A recent field experiment in a supermarket chain shows that this interaction between bonuses and feedback may be even more intricate [9]. The study randomly allocated 224 supermarkets to three different treatment groups and a control group. One treatment group received a monetary bonus to increase store profits. In another group, no bonus

was paid but store managers had biweekly performance review conversations about what they did to increase profits. Finally, the third treatment group received both a bonus and had the review conversations. The review conversations raised profits by about 8%. The bonus, however, had no discernible effect on performance. Moreover, the bonus even undermined the value of the review conversations, as performance in the combined treatment also did not exceed performance in the control group. Evidence from analysis of protocols of the review meetings shows that the bonus payment changed the nature of the review conversations. For instance, store managers spoke significantly less often of encountered problems when the bonus was in place and perceived feedback quality was higher when no bonus was in place. The authors argue that performance review conversations trigger reputational incentives as store managers want to signal their motivation to supervisors and these reputational incentives are undermined when the bonus is in place. An implication for the design of bonus plans is thus to carefully evaluate their interplay with other management practices.

Subjective performance evaluations

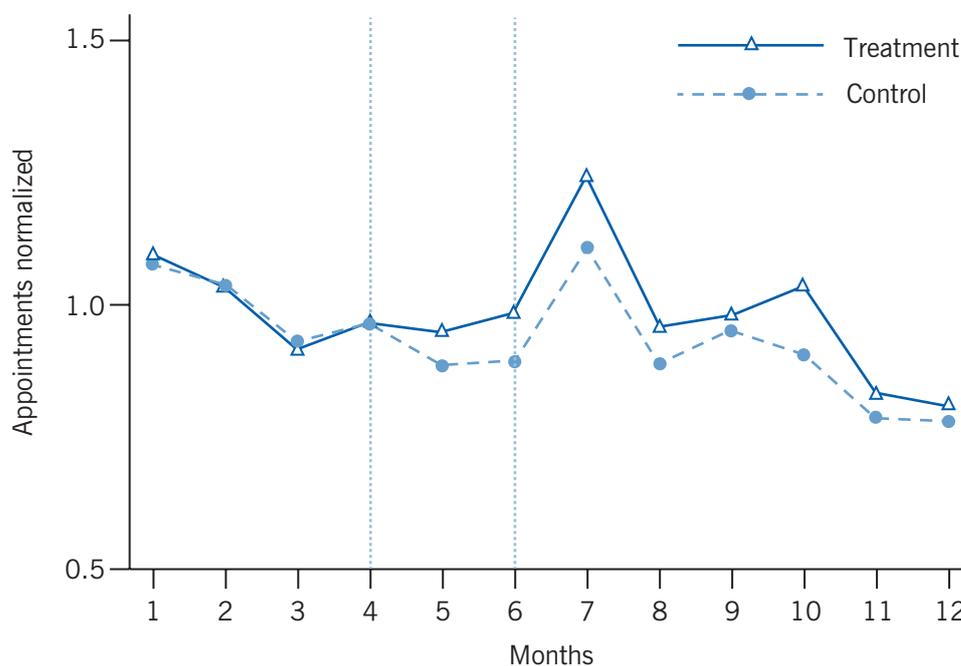
When objective performance measures are unavailable, firms have to rely on subjective performance evaluations when implementing bonus plans. A typical procedure is that bonus payments are determined by performance appraisals in which an employee's performance is assessed on a rating scale (most commonly by their supervisors). Common rating scales include five-point scales, where for instance a rating of one indicates the highest and a rating of five the lowest performance. Other appraisal formats follow a management-by-objectives (MbO) approach where supervisors set targets and assess an employee's target achievement (typically as a percentage of the target, such that a rating of 100% would mean that the employee would have achieved all targets and a rating larger than 100% that the employee exceeded their targets). Finally, some firms, particularly those in banking and finance, use bonus pool arrangements, where the financial performance of a unit determines the bonus pool amount and a supervisor subjectively decides how to allocate this pool to their subordinates. In all these schemes supervisors have some discretion in how bonuses are distributed as true performance is not objectively verifiable. A large literature in psychology and economics has established that these subjective ratings tend to be biased. A typical claim is that ratings tend to differentiate too little between high and low performance ("rating compression") and that—when there is no fixed budget—ratings tend to be too generous toward employees ("leniency bias"). Core reasons for such distortions encompass restricted observability (limiting the ability of supervisors to differentiate), and social preferences (triggering more generous ratings). Moreover, a substantial body of evidence has shown that people tend to evaluate wages (and bonuses) relative to specific reference points such as the income of colleagues. When supervisors anticipate this, it may again lead to a reluctance to differentiate between high and low performers. Such biases may undermine the effectiveness of bonus schemes that rely on subjective assessments.

A study on the introduction of objective key figures in a retail bank provides causal evidence for the claim that subjective assessments are biased and that these distortions restrict the benefits of performance bonuses [10]. Prior to the intervention, the bank used a bonus pool arrangement such that branch managers had to allocate bonuses based on their subjective assessments. The bank then ran a field experiment, providing precise objective

sales figures to branch managers in a randomly selected subset of all branches (from the sales IT system) which supervisors then used to allocate the bonus pool.

The introduction of objective performance measures indeed increased employee efforts (as measured by employee-initiated customer appointments) and profits. Figure 1 shows the development of employee-initiated customer appointments over time. In the treatment group it was announced after month 4 that objective performance measures would be available starting with month 7. The analysis also reveals that performance gains were particularly large in larger branches, indicating that subjective assessments are less accurate and thus objective performance information more useful when supervisors assess more employees at the same time.

Figure 1. Bonuses tied to objective metrics improve employee performance



Notes: The figure shows the number of employee-initiated customer appointments (normalized at the average of the four months prior to the intervention—in month 7) over 12 months in the treatment and control groups.

Source: Manthei, K., and D. Sliwka. “Multitasking and subjective performance evaluations: Theory and evidence from a field experiment in a bank.” *Management Science* 65:12 (2019): 5861–5883 [10].

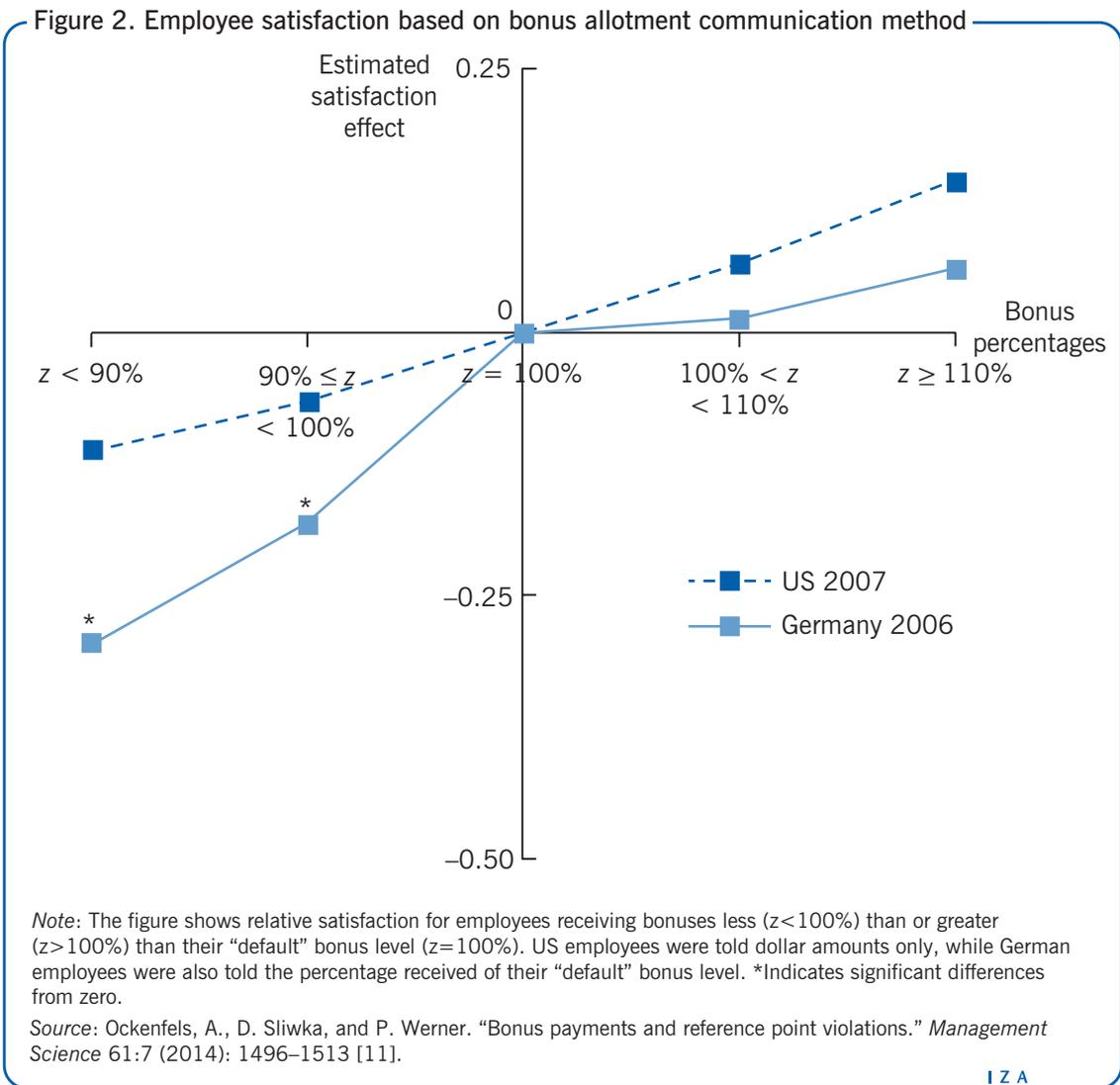
I Z A
World of Labor

The combination of personnel and survey data allows a deeper understanding of behavioral drivers of bias in subjective assessments. A field study in a multinational company provides evidence for the specific role of reference points and equity concerns in evaluating bonuses [11]. The company used a bonus scheme in which supervisors appraised their subordinates’ performance subjectively and then assigned an annual bonus. For each employee there was a “bonus budget” (determined by company and sub-unit performance) and supervisors could reallocate this budget among their subordinates. For each subordinate they had to determine a “bonus percentage,” that is, the percentage share of the budget assigned to this employee. Receiving a bonus percentage below 100% thus implied that an employee received a lower bonus than the average of their colleagues. The firm applied the scheme for their managers in Germany and the US—with only one

difference: appraised employees in Germany learned their exact bonus percentages but employees in the US only learned the dollar amounts of the bonus.

The study then investigated the association between assigned bonus percentages (containing information about an employee’s relative standing) and job satisfaction. Figure 2 shows this association, revealing that employees in Germany who learned that they received less than 100% of their allotted bonus budget (and thus less than their colleagues on average) were significantly less satisfied, while those receiving more than 100% only barely exhibited increased satisfaction. Interestingly, this pattern was absent in the US. The 100% bonus level thus constituted a clear reference point for comparison when known to employees. It was furthermore observed that managers in Germany apparently tried to avoid such “reference point violations” by compressing bonus payments. The fraction of German employees with an evaluation at exactly 100% was more than twice as high as in the US, supporting the view that equity concerns can be key drivers of rating compression.

This finding raises the follow-up question of whether rating compression undermines performance or whether it may also have beneficial effects. For instance, a lack of differentiation in evaluations may lead to higher overall job satisfaction which in turn



can improve employee motivation. Alternatively, overly high levels of differentiation may undermine employee willingness to cooperate with colleagues which may harm performance. This issue has been empirically explored in a study using panel data from a Swiss unit of an international company, showing that more variability in bonus payments is positively associated with more overtime work [12].

A related study explores the connection between dispersion in bonus payments and the size of subsequent bonus pools in a panel of personnel data spanning a large number of German banks [13]. This study generally finds evidence in line with the hypothesis that differentiation increases subsequent financial performance. The authors also ran a survey among experts in banking asking them to rate different functions according to the extent to which they believe that individual performance can be assessed objectively. Bonus differentiation was found to be more valuable in functions where this is the case (such as retail and investment banking and asset management rather than corporate banking and back office functions). Interestingly, the association between dispersion and subsequent performance was stronger at higher hierarchical levels but may actually be reversed at the lowest levels—indicating that equity concerns and cooperation may be more important for performance at lower hierarchical levels.

Firms that want to foster differentiation in evaluations sometimes implement forced distributions—that is, they introduce the obligation that managers have to follow a specific distribution in their rating (thus, for instance, assigning low ratings to a sufficiently high fraction of employees). While there is evidence from lab experiments showing that forced distributions raise performance if workers work independently and reduce performance if they can harm each other, there is as of yet no firm-level field experiment on the costs and benefits of forced distributions.

LIMITATIONS AND GAPS

A growing number of field experiments are being conducted in firms to estimate the performance effects of bonus schemes. However, each field experiment studies a specific firm and often even a specific job type. And, as this article has argued, context matters: Which performance measures are available? What are the task preferences and personality traits of the employees? What other management practices are in place that affect performance and how do these practices interact with bonus payments? Relative to the aim of obtaining a comprehensive picture of these interdependencies, the number of existing field experiments on bonus design in firms is still rather small. Each firm that evaluates its compensation scheme via a field experiment not only learns about optimal compensation in its own environment, but also contributes to completing this picture.

SUMMARY AND POLICY ADVICE

Bonus plans can work, but they are no panacea to motivate employees in firms. When simple key performance measures are available that capture key elements of an employee's work and when they are hard to monitor, then bonus payments based on individual employee performance measures tend to work as expected. However, individual performance bonuses may have limited effects for certain types of employees, for instance those who are strongly self-motivated to give their best. Furthermore, other management practices such as symbolic rewards or performance review conversations can alternatively motivate

employees. Moreover, when individual performance cannot be assessed objectively, subjective assessments must be used to determine bonuses and these tend to be biased. An alternative to individual performance bonuses are team bonuses based on the success of a specific team or of the whole organization. While economists have typically feared free-rider problems, there is mounting evidence that team bonuses may work quite well as they raise both incentives and cooperation.

It is clear that context matters when judging whether and what type of bonus may be able to raise performance in a specific environment. The design of incentive schemes should thus be viewed like an engineering task: it starts with a detailed analysis of the considered jobs and previous evidence. Based on this analysis a specific bonus plan design is developed. Given the complexity of the task, sound evidence-based bonus design then aims at using its implementation within the organization to evaluate its effects. And a proper evaluation only works when the design is implemented first for a (ideally randomly selected) subset of the organizational units. By comparing outcomes between subgroups such evidence-based bonus design helps to verifiably identify the effects on performance and employee well-being.

Acknowledgments

The author thanks an anonymous referee and the IZA World of Labor editors for many helpful suggestions on earlier drafts. Previous work of the author contains a larger number of background references for the material presented here and has been used intensively in all major parts of this article [9], [10], [11], [13].

Competing interests

The IZA World of Labor project is committed to the IZA Code of Conduct. The author declares to have observed the principles outlined in the code.

© Dirk Sliwka

REFERENCES

Further reading

DeVaro, J. *Strategic Compensation and Talent Management: Lessons for Managers*. Cambridge: Cambridge University Press, 2020.

Prendergast, C. “The provision of incentives in firms.” *Journal of Economic Literature* 37:1 (1999): 7–63.

Key references

- [1] Lazear, E. P. “Performance pay and productivity.” *American Economic Review* 90:5 (2000): 1346–1361.
- [2] Hong, F., T. Hossain, J. A. List, and M. Tanaka. “Testing the theory of multitasking: Evidence from a natural field experiment in Chinese factories.” *International Economic Review* 59:2 (2018): 511–536.
- [3] Bandiera, O., I. Barankay, and I. Rasul. “Social preferences and the response to incentives: Evidence from personnel data.” *The Quarterly Journal of Economics* 120:3 (2005): 917–962.
- [4] Donato, K., G. Miller, M. Mohanan, Y. Truskinovsky, and M. Vera-Hernandez. “Personality traits and performance contracts: Evidence from a field experiment among maternity care providers in India.” *American Economic Review Papers and Proceedings* 107:5 (2017): 506–510.
- [5] Knez, M., and D. Simester. “Firm-wide incentives and mutual monitoring at Continental Airlines.” *Journal of Labor Economics* 19:4 (2001): 743–772.
- [6] Hamilton, B. H., J. A. Nickerson, and H. Owan. “Team incentives and worker heterogeneity: An empirical analysis of the impact of teams on productivity and participation.” *Journal of Political Economy* 111:3 (2003): 465–497.
- [7] Friebe, G., M. Heinz, M. Krueger, and N. Zubanov. “Team incentives and performance: Evidence from a retail chain.” *American Economic Review* 107:8 (2017): 2168–2203.
- [8] Delfgaauw, J., R. Dur, J. Sol, and W. Verbeke. “Tournament incentives in the field: Gender differences in the workplace.” *Journal of Labor Economics* 31:2 (2013): 305–326.
- [9] Manthei, K., D. Sliwka, and T. Vogelsang. *Talking about Performance or Paying for it? Evidence from a Field Experiment*. IZA Discussion Paper No. 12446, 2019.
- [10] Manthei, K., and D. Sliwka. “Multitasking and subjective performance evaluations: Theory and evidence from a field experiment in a bank.” *Management Science* 65:12 (2019): 5861–5883.
- [11] Ockenfels, A., D. Sliwka, and P. Werner. “Bonus payments and reference point violations.” *Management Science* 61:7 (2014): 1496–1513.
- [12] Engellandt, A., and R. T. Riphahn. “Evidence on incentive effects of subjective performance evaluations.” *ILR Review* 64:2 (2011): 241–257.
- [13] Kampkötter, P., and D. Sliwka. “More dispersion, higher bonuses? On differentiation in subjective performance evaluations.” *Journal of Labor Economics* 36:2 (2018): 511–549.

Online extras

The **full reference list** for this article is available from:

<https://wol.iza.org/articles/bonuses-and-performance-evaluations>

View the **evidence map** for this article:

<https://wol.iza.org/articles/bonuses-and-performance-evaluations/map>