NBER WORKING PAPER SERIES

WHAT DO PERFORMANCE APPRAISALS DO?

Peter Cappelli
Martin Conyon

Working Paper 22400
http://www.nber.org/papers/w22400

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
July 2016

We are grateful to seminar participants at University of Maryland, MIT, and the University of Pennsylvania for helpful comments. We especially thank David Almeda for discussions and access to the data used in this study. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

What Do Performance Appraisals Do?
Peter Cappelli and Martin Conyon
NBER Working Paper No. 22400
July 2016
JEL No. J33,J41,J63

ABSTRACT

This paper investigates employee performance appraisals using data from a single US firm between 2001 and 2007. We find that performance appraisals are both informative and drive important components of the employment contract. We find that employee appraisal scores vary considerably both between and within individuals over time. In addition, we show that employee performance appraisal scores are related to a range of important employment outcomes, including merit pay and bonuses, promotions, demotions and dismissals, as well as employee quits.

Peter Cappelli
The Wharton School
Center for Human Resources
University of Pennsylvania
Philadelphia, PA 19104-6358
and NBER
cappelli@wharton.upenn.edu

Martin Conyon
Senior Fellow
Center for Human Resources
The Wharton School
University of Pennsylvania
Philadelphia, PA 19194-6358
conyon@wharton.upenn.edu

What do Performance Appraisals Do?

Introduction

Performance appraisals are the process through which supervisors assess, after the fact, the job-related performance of their supervisees and allocate rewards to the supervisees based on that assessment. They are an essential part of the effort to address agency problems, getting employees to act in the interests of the employer, and, as such, are a central practice in the field of management. How they operate, and whether they are effective, are also the subject of an academic debate: Are performance appraisals best seen as a component of simple contractual relationships or as a part of a richer relational contract between organizations and employees?

Our study provides important new evidence on how performance appraisals work at a major US corporation. We show first that such appraisals are informative to management, in contrast to some critiques. Second, the evidence is not consistent with idea that they operate as part of a contractual relationship. Performance appraisals seem instead to support relational contracts between the employees and the firm.

The performance appraisal has been traced back to the early 1800s and the UK cotton mills owned by Robert Owen. They became popular in the US after WWII (see Murphy and Cleveland 1995 for a survey). They are now so common in the US and apparently so taken for granted that there have been almost no attempts to establish how extensively performance appraisals are used. One of the few exceptions was a small 1996 survey by the Society of Human Resources Management, which found that all but 2.8 percent of US respondents reported that their organization had a performance appraisal process (those reporting otherwise include very small operations), a rate of use that has increased over time (see Roslyn 1996). A more recent survey (Aberdeen Group 2010) put the figure at 91 percent for employers around the world. Performance appraisals are required in 75 percent of individual jobs in US state government (Seldon, et al. 2001), most aspects of the Federal government, in the military, and indeed in virtually every type of employment relationship. Important exceptions include most unionized operations and many - but not all - professorial jobs.

Despite the near universality of performance appraisals, little is known about the most basic questions concerning them, such as why they exist and how they are actually used (see Milkovich and Wigdor 1991 for information on their use in the US in an earlier period). Commenting on this, Prendergrast (1999. p.33) notes: “A disappointment of the economics literature has been the paucity of information collected on the evaluation of workers with poorly measured output,” the arena that performance appraisals target. Borjas’ (2010) encyclopedic survey of economic research on labor and employment topics, for example, shows no studies on

performance appraisals. As Indjejikian (1999) observes, the lack of study about what appraisals actually do is especially unfortunate because of the central role they play in addressing agency problems, perhaps the main issue in managing firms.

There is an extensive literature about performance appraisals in industrial psychology. Murphy and Cleveland (1995) survey hundreds of such studies. As they note, these studies view the important issues in appraisals as measurement problems: How to design scales for the appraisal form, how to reduce individual biases, the effect of individual attributes on the process, and so forth. The reason for the existence of performance appraisals is rarely if at all considered. Their survey notes no studies that examine the effects of appraisal scores on employment outcomes, the central concern here. In another review, Ilgen, Barbes-Farrell, and McKellin (1993) conclude that the main focus of this research had been on assessing the attributes of the raters. A more recent review by Picher (2012) suggests that perceived acceptance of the process by employees is arguably the central interest in these studies. Longenecker et al (2014) note the relative absence of even descriptive data on actual performance appraisal practices.

In 2016, we examined 400+ recent, peer-reviewed articles in the encyclopedic EBSCO database that contained the phrase "performance appraisal" in either their title or their abstract. We cite all those that address employment outcomes associated with performance appraisals below.

One debate about performance appraisals, described in more detail below, stems from practitioner complaints that appraisals do not in fact differentiate in any serious way across employees. Practitioners and academics have raised similar concerns as to whether the appraisals have much influence on employee outcomes, such as pay increases and promotions. We examine that view empirically.

The most common academic view of appraisals tends to be that they are an adaptation to the limits associated with explicit contracts in the workplace. Often they are seen as the equivalent of a post-contract review, settling up an employee's compensation based on their performance over the previous period and as a supplement to explicitly contractual arrangements, addressing difficult-to-measure aspects of performance and using discretionary compensation to reward those aspects.

An alternative view is that performance appraisals are a central component in what we might describe as a relational contract, where employment is seen as an open-ended employment

relationship moderated by supervisors. We examine these two competing views regarding employee performance appraisals below.

Theoretical Considerations

If we think of the employment context as creating an agency problem - how to get employees to act in the interests of their employer – the typical solution in standard agency models is some form of contract specifying rewards contingent on performance: Outline the desired outcomes, see if they are achieved, and reward employees accordingly. By that view, performance appraisals are simply a settling up of the prior agreement linking incentives to job performance.

A caveat to this view is that many and arguably most jobs are too complex to be governed by simple contracts that require performance to be defined and measured in simple ways (Lazear (1986, 2000).[1] As Baker (1992) and Baker et al (1994a, b) note, performance appraisals provide a means for assessing the performance of tasks that may have no easy, objective performance metrics, such as the quality of employee interactions with customers or peers. Hence models draw a distinction between objective measures of performance (those that are at least conceptually straight-forward to measure) and subjective measures (those that are conceptually difficult to measure), where performance appraisals assess the latter. See MacLeod (2003) for a formal model.

The more subjective aspects of performance there are, the more we should see mechanisms like performance appraisals (MacLeod and Parent 1999). As Brown and Heywood (2005) find with Australian data, jobs that are more complex and have more separate tasks are more likely to be covered by performance appraisal systems and associated merit pay arrangements. Evidence suggests that simply having performance appraisals is associated with higher levels of individual performance (Marc-Arthur et al. 2009) as compared to individuals in jobs that do not have appraisals, presumably because of this ability to capture and manage the subjective aspects of performance. Prendergast and Topel (1996) outlined the purposes that such subjective assessments of employee performance can serve.

Anecdotal evidence supports the importance of subjective aspects of performance even in jobs where one would think performance a priori would be highly objective. Financial traders who manage portfolios, for example, have objective and straight-forward measures of their performance: the financial returns on their portfolio. But even in this context, their employers

[1] One may think of the job, J, as being made up a set of tasks, T_i where i=1,…,t. If one or more of the tasks is omitted, J is measured with error. The overall effect of this on employee behavior may be quite large.

retain a subjective performance appraisal system, asserting that it is necessary because traders have many ways in which they can help or hurt the firm through behavior that is not captured by the returns they generate from their portfolio (Fenton-O'Creavy, et al. 2007). A similar example comes from car sales where subjective appraisal elements are also used to reward sales representatives, apparently because the value of cars sold captures only one relevant aspect of performance (Gibbs, et al. 2004).

The notion that there can be explicit contracts and subjective elements to them has been considered formally at length (e.g., Levin 2003; Rayo 2007). Special attention has been given to examining those subjective aspects and especially the associated discretionary rewards (Baker, Gibbons, and Murphy 1994; Gibbs et al 1994; Yang 2008). Indjejikian and Mateika (2012) find that business unit managers appear to influence internal accounting practices so that subjective appraisal measures recognize performance factors that are important to them. Arya and Mittendorf (2011) model the conflict between short-term performance and long-term career aspirations and how subjective performance measures can resolve it.

A different view than the notion of performance appraisals as either an explicit or supplemental part of a contractual arrangement begins with the concept of employment itself as an informal, relational contract. The recognition that many economic exchange relationships are governed by something other than explicit contracts has a long history in fields like anthropology and especially sociology, where it is one of the fundamental topics (e.g., Macauley 1963). The concept of relational contracts has been used to explain an array of practices in the economy, from arrangements governing contractors in the construction industry (Cheung 2006) to strategic alliances set up by venture firms (Lindsey 2008), to the historical legal case concerning the relationship between Standard Oil and the Pennsylvania Railroad (Brown 2011), to the retail price recommendations that manufacturers issue for sellers (Bueller and Gartner 2013).

While there are potentially many kinds of such relationships, including reciprocity-based gift exchanges, the notion of exchange relationships that are governed by self-interest became known as relational contracts. These are perhaps most easily defined by what they are not: An informal (as opposed to written) understanding that is self-enforcing (as opposed to being enforced by the courts or a third party) that governs economic activity between parties. A key component of relational contracts is that the parties have an interest in them continuing so that each side has an interest in addressing imbalances in contributions and rewards that might otherwise cause the arrangement to end. An advantage of these relational contracts is their presumed flexibility. They can accommodate changing circumstances quickly, without the need

for redrafting a formal agreement, they can handle situations where outcomes cannot easily be measured, and they also avoid the slow and costly enforcement process of the courts. There are now over 750 academic studies as measured by the EBSCO database across fields that make use of the idea of relational contracts.

Employment as a relational contract: By far the most important context for these ideas is employment. Contemporary arguments about relational arrangements in employment might be said to begin with Simon's (1951) exploration as to how employment is different from sales contracts, focusing on the appeal to long-run interests as a means for addressing the incentives to cheat. Klein, Crawford, and Alchain (1978) is among of set of papers that broadened the discussion, observing the range of relationships in business that were not captured by contracts nor vertical integration. Williamson (1985) brought the concept of relational contract to significantly greater attention, on par with markets and hierarchies as an organizing principle. Since then, much of the attention has understandably focused on the enforcement mechanisms that allow these arrangements to persist, with game-theoretic explanations based on long-run interests in continuing relations being the most important explanation (Bull 1987, Baker, Gibbons, and Murphy 2002).

The reason why employment has been seen as a relational contract has to do with the reality of employment itself. Employment relationships, at least in countries governed by common law legal systems, can be thought of as incomplete and open-ended contracts. Employers are allowed to adjust the relationship unilaterally and at any time, changing the tasks they want employees to perform and changing the rewards associated with those tasks.[2] Ups and downs in customer demands, turnover of other employees, use of new technology, or even changes in business strategy (e.g., product mix) can change what tasks employees perform and how they execute them. As Brickley, Zimmerman, and Smith (2008 Chapter 16) note, employers can change performance standards for individuals as they learn more about the person's capabilities: The better one performs, the more the employer expects of them. Unlike in

[2] In the US, the "at will" employment notion underpins employment relationships and gives employers virtually unlimited authority to direct and manage employees. State-level restrictions on the rights of employers (e.g., prohibiting employers from requiring employees to commit crimes) pare back but in no way eliminate the basic right of employers to change job and performance requirements unilaterally. Even in unionized operations, employees are required to follow the directions of employers and only after the fact challenge those directions through the grievance procedure if they violate collective bargaining agreements. See Attelson (1978) for the legal principles underpinning employment relationships. Brodie (2011) discusses the UK context.

contractual arrangements, employers do not need agreement from employees to make such changes.

Not only can employers change the tasks employees perform, how they perform them, and the standards for performance during the accounting period, they are also allowed to change arrangements after-the-fact. Employers can change the criteria for and the application of pay increases, bonus payments, and other rewards based on their business conditions or other circumstances. This distinction is consistent with the Baker, Gibbons, and Murphy (1994) observation that employers can use the appraisal process in part to "back out" changes in context that might make some employee goals easier and others harder to achieve than anticipated.

Unlike post-contract reviews in particular, performance appraisals are forward-looking with respect to their consequences and one-sided. A post-contract review between a homeowner and a builder, for example, could lead to a transfer of assets from the builder to the homeowner (e.g., because of failing to meet performance standards) or from the homeowner to the builder (e.g., because of cost overruns) in order to settle-up the original contract. There is no equivalent in the performance appraisal process. There is no attempt to "claw back" resources from employees who under-performed. Wages are owed to employees just for working and cannot be made subject to after-the-fact performance appraisal assessments, at least for employees covered by the Federal and State wage and hour laws. Even where such laws do not apply, claw back arrangements for salaries are practically unheard of. Rewards for good performance and punishments for poor performance affect only the next period of the relationship. (Dismissal, which is the most severe punishment, ends the relationship in the next period.)

Other evidence suggests that the performance appraisal process cannot be even a supplement to the basic contracting model. As Hart (1995) observes, the possibility of legal enforcement is central to the notion of contractual aspects of firm. Performance appraisals, however, cannot be adjudicated by the courts, at least in the U.S.[3] This distinction also explains

[3] See Krenik v. County of Le Sueur, 47F.3rd 953, 960 (8th Cir.1995), affirming that the courts defer to the employer on most personnel matters, including performance appraisal outcomes. The courts clearly can and do intervene when the performance appraisal process interferes with other legal standards, e.g., if it contributes to discrimination. In terms of their effects, performance appraisals are judged as a type of employment test subject to the Uniform Guidelines on Employee Selection created by the Equal Employment Opportunities Commission in 1978. Under those guidelines, performance appraisals if used as the basis for promotions are required to demonstrate validity in the sense of being accurate descriptors of past performance and significant predictors of future performance (see Murphy and Cleveland 1995 for a discussion). Otherwise, as long as performance appraisals are not attached to a contractual agreement that goes beyond standard employment relationships, the courts do not have jurisdiction over them.

why performance appraisals are rare in union settings, because formal contracts do govern the employment process in that context.

The idea that performance appraisals simply capture the subjective aspects of performance, adding that information to the more objective performance measures that are governed by something closer to a real contract, is also contradicted in practice. Virtually all appraisals combine objective, readily available measures of performance, such as absenteeism, with subjective measures into a single variable without using consistent formulas or weights (see the review by Murphy and Cleveland 1995). The supervisor then uses their own judgment as to how to produce the overall score. This process would appear to negate the advantages of validity and reliability provided by the objective information if the goal was only to enforce an agreed upon contract linking rewards to performance.

Performance appraisals are typically conducted by the immediate supervisor. While it is widely recognized in the academic literature that supervisors monitor job performance, it is less well acknowledged that they also direct employees as to which tasks to perform and how to carry them out. (Among the many descriptions of supervisory responsibilities, which indicate that the monitoring function is a minor task, is Mosely et al. 2010.) As observers of performance appraisals in firms note, the individual agendas of supervisors are therefore an explicit part of the appraisal process (Longenecker et al. 1987). The growing interest in seeing how attributes of supervisors might influence the appraisal scores they give reflects this orientation (e.g., Bernardin et al 2016; Harari et al. 2015).

A consequence of the above is that the supervisor is *defining* what constitutes good performance *and also assessing* whether it was achieved. By analogy, this is the equivalent of a referee who decides what the rules of the game should be as the game goes along and then also enforces those rules. While this arrangement may seem strange, it is a necessary adaptation to the open-ended, common law approach to employment, specifically the right to redirect employee efforts unilaterally. Because the supervisor has the ability to use their own judgment to tell workers what to do and how to do it, changing those directions as circumstances dictate, the supervisor is virtually the only person who could assess the performance of those workers.

The above distinction helps explain why performance appraisals are so much more prevalent in common law countries, where the open-ended, relational concept of employment applies. (See Bloom and Van Reenen 2010 and Peretz and Fried (2012) for some cross-national evidence on performance appraisals.) This relational view of appraisals also has implications for estimation. Identical employee behavior in identical jobs for employee *a* and *b* could be

objectively rated as different if the supervisor over the course of the accounting period instructed those two employees to emphasize different tasks within their jobs or held different expectations for them based on different abilities.[4] Variation in appraisal scores for similar jobs across supervisors, indeed within supervisors, may well be informative rather than the result of assessment error.

The principal of 'informativeness' as a signal of the employee's effort (Holmstrom, 1979) suggests that any variable that increases the accuracy with which worker performance can be measured should be included in an assessment. Yet the supervisor creates his/her own index from the separate and typically disparate measures. The reason is, given the distinctions above, only the supervisor is in a position to know what individual employees were directed to do over time, the priorities and expectations held for various tasks, and what standards constitute good results. Even "objective" measures such as sales per person may not be accurate measures of individual performance if, e.g., the supervisor instructed the employee to take time away from selling to train a subordinate or if business conditions deteriorate. Only the supervisor is in a position to assess performance accurately. If performance appraisals were simply part of enforcing a contract, then the common practice of combining objective measures and subjective measures into a single variable would defeat the purpose by adding noise to the measure.

If performance appraisals were just part of a post-contract review, employers would appear to have an incentive to understate performance and associated rewards in such a model. Indeed, eliminating these individual appraisals and associated rewards is a common goal of unionization, and greater transparency in contractual arrangements and on performance might reduce the ability of employers to cheat employees. The fact that both parties usually have an incentive to continue the relationship provides the best explanation as to why cheating is not widespread. The appraisal-based arrangement can be seen a repeated game that can be self-enforcing if the two parties have compatible expectations. In the employment context, the fixed costs of employment reduce the incentive for employers to risk employee quits because of

[4] Careful readers will see that this discussion could quickly take one into the topic of epistemology and meaning more generally. The distinction between the "positivist" position, which holds that the notion of objective reality is overrated, and the "realist" position, which asserts that it is not, is an obvious extension. This distinction is especially important in law (see Green 2003): Are the interpretations of the US Constitution by the Supreme Court, for example, in fact the real meaning of the Constitution, as legal positivists might claim? Or is it possible to say that there is an objective meaning to the Constitution and that the Supreme Court can make mistakes in its interpretation, as realists claim? We are not making a positivist claim about supervisor-based appraisals. We are instead simply claiming that the supervisor as a condition of the employment relationship gets to make the rules as they go along. It is still possible for the supervisor to make errors in judgment about applying those rules.

cheating, and because the consequences of appraisals apply only in the next period, employees have the option of quitting before experiencing the consequences of a fraudulent appraisal.

What Appraisals Do in Practice

Despite the conceptual arguments above suggesting how functional performance appraisals should be, in practice they are one of the most unpopular and criticized aspects of the modern workplace. Critiques of the performance appraisal process date back at least to Likert (1959) and have appeared continuously since. Davis and Landa (1999) present survey data suggesting that the practices are dysfunctional and Cutler (2008) argues that the entire process should be abandoned.

Perhaps the most fundamental criticism of appraisals is the claim that supervisors are reluctant to differentiate employees based on performance because doing so creates hard feelings and divisions in the workforce. Prendergast (1999) reports early evidence of this but notes the absence of contemporary tests of this claim. (See Mani (2002) for a public sector illustration of this bias.). The concern about the reluctance to differentiate performance has also been modeled by Prendergast and Topel (1996), who suggest a solution to it.

A more modest version of this basic complaint is that supervisors are reluctant to give poor ratings (see, e.g., Thompson and Radin 1997 and Medoff and Abraham 1980 for empirical evidence). Under this view, they may differentiate among above average employees but resist giving below average scores. This should lead to inflated average scores and a truncated distribution of scores associated with left-censoring of the distribution, what we might call the "Lake Wobegon" effect, where every employee is above at least the average point on the scale. Grund and Przemeck (2012: abstract) reflect this prevailing wisdom (arguably overstating it) in their modeling of appraisals by observing that "it is a well-established fact that many supervisors tend to assess the employees too good (leniency bias) and that the appraisals hardly vary across employees of a certain supervisor (centrality bias)." Evidence suggests that procedures for getting supervisors to take the subjective decisions in performance appraisals seriously can improve their validity (Meno et al. 2003).

A related version of the lack of differentiation complaint is that once supervisors get to know employees and develop personal relationships with them, they are reluctant to give them tough evaluations (see also Milgrom (1988) and Milgrom and Roberts (1988). When supervisors change and those relationships are broken, the argument goes, we should see tougher appraisal scores.

A final version of the notion that performance appraisals are not functional is based on the assertion that job performance is dispositional, based on underlying aspects of initial human capital when entering the firm that cannot be changed. Beyond the initial assessment of employee performance, therefore, continuing appraisals add little information or value. The popular version of this argument is associated with the "War for Talent" (Michaels, et al. 2001) notion that there are good workers and bad workers, with the implication that the employer's job is simply to screen and sort: Keep the good workers, and get rid of the bad ones. If this version of the complaint is true, we should see little or no change in the scores of individuals over time.

More generally, the extent to which performance is a function of initial attributes and human capital as opposed to how individuals are managed within the firm is an open question for which there has been little evidence. The assertions above have not been examined in any serious fashion. We address them below.

The most important research questions about performance appraisals concern their relationship with employment outcomes. While actual performance appraisals may have other uses as well, the main justification for having them is to motivate employees to perform their jobs well. Some part of that influence comes from the feedback about where performance was good or not so good and what to do differently going forward. An important influence also comes from the incentives they create to perform well, which in turn are based on how rewards and consequences are tied to performance appraisal outcomes.

The contractual, settling-up view of performance appraisals implies that the objective in designing appraisals should be to make as close, consistent, and predictable a connection between job performance in the previous period and consequences in the next period as possible: Ideally, rewards in $t+1$ should be perfectly correlated with performance in t as motivation declines when the connection between performance and consequences is weakened and uncertain.

The relational view of employment suggests different objectives and different practices. Especially where the presumption is that employment will continue indefinitely, employees represent something of a fixed cost. We therefore might expect supervisors to try to improve employee performance rather than just dismiss those with lower performance. Specifically, we expect employers to recognize and reward *improvements* in performance explicitly, independent of the level of performance. The contractual view of the appraisal, in contrast, rewards only level of performance. This difference offers the opportunity for an empirical test of which of these two views is a more accurate description of what performance appraisals do and is the main hypothesis we examine below.

We also might expect supervisors in a relational contract to use some discretion in rewarding performance to build relationships and encourage future performance. For example, supervisors may disproportionately reward higher levels of performance in order to reinforce in a stronger way the motivation for higher performance in the next period. Disproportionate rewards that were not anticipated would be inconsistent with the settling up notion of contracts. Baker et al (1994, 1994b) find evidence that the relationship between job performance and rewards was not consistent across pay grades and organizational levels, suggesting that those relationships could not have been anticipated and would therefore be inconsistent with a settling-up view of appraisals. In their context, however, the variation was more likely attributable to the constraints of compensation systems rather than the discretion of supervisors.

Arguably the most important reward has been pay increases in the next period or merit pay. Annual bonuses are also thought to vary with job performance over the previous accounting period. It is common to believe that promotions, demotions, and dismissals are related to performance appraisal outcomes as well. There is little direct evidence as to whether those pay and employment outcomes actually vary with appraisal scores, however. Exceptions include Medoff and Abraham's (1980, 1980b) conclusion that performance appraisal scores did not drive wage increases over time in the organization they studied and that tenure was the key factor. Gibbs and Hendricks (2004) study of pay practices in a large firm, on the other hand, concluded that appraisal scores were associated with higher merit pay and bonuses. We reexamine these conclusions directly.

We first use descriptive data to assess the extent to which performance appraisal results are informative, testing assertions from practitioners that the performance appraisal process is not functional. We then use within-person fixed effects estimates to examine the extent to which appraisal outcomes are dispositional.

Most importantly, we test whether the appraisal process is best seen as part of a contractual relationship or as part of a component of an open-ended, employment relationship. Evidence that differentiates between these will be whether supervisors reward improvements in performance, whether performance is rewarded through long-term rewards, such as promotions, and whether supervisors manage the relationship between prior performance and financial in ways that depart from linearity.

Data

We examine the above questions using unique panel data on managerial employees from a single large, publicly traded US firm. The company analyzed in this study is a retail business in the

S&P500 index with over 50,000 employees. A small number of observations from outside the US are excluded because regulatory environments and other differences in context may cause practices and relationships to differ in spurious ways. The fact that we have these data over multiple years allows us to control for person-specific effects, a crucial advantage in assessing the above arguments.

Using data from a single firm obviously controls for cross-firm heterogeneity in the appraisal process that might confound the above relationships, albeit at the cost of generalizability. We do not claim that this firm is representative of a typical US employer. Retail operations employ about 11 percent of US workers[5], however, and the nature of jobs in this company appears to be no different than in most retail operations: There is no significant proprietary technology or any proprietary products in this company, as is the case with most retail operations. Although the company is much larger than typical US employers, the individual stores, which account for virtually all its jobs, are no different from those in most other retail operations. In short, we believe these data are representative of retail operations per se, especially at the store manager level, which constitutes the vast majority of our observations.

We have information on all managerial employees in this company over the period 2001 to 2007, which includes their performance appraisal scores, a series of employment outcomes, and demographic attributes. The lowest-level managers, which account for 96 percent of the observations, are store managers. An advantage of examining appraisals for this group is that there is considerable scope for variation in performance as the jobs are reasonably complex. But unlike higher-level managers, more aspects of performance at the store-level are within the control of the individual manager. The Bureau of Labor Statistics reports that "management, professional, and related occupations" account for about 40 percent of all jobs in the US (BLS 2014). Lower-level managers, like the ones in our sample, are certainly the vast majority of that category. We observe individuals from the lowest level, assigned a value zero, to the highest level assigned a value of ten. In the regressions below, the excluded category is the lowest level (i.e. job level=0) so the job level effects are measured relative to them.

There were 23,427 person-year records over the sample period and 4,568 individual-level records included in the input dataset, representing over 98 percent of the active individuals in the management population in this firm. The most senior executives, designated as 16(b) officers, were excluded because they do not appear to have performance appraisals. There were five such

[5] See http://www.bls.gov/opub/ee/empearn201008.pdf.

executives per year on average designated as 16(b) officers in the company.[6] We also exclude managers employed in other countries where different practices may apply. Our analyses are based on results for 4408 managers.

Performance appraisal measure: The performance appraisal process in this company is very similar to practices we believe are common in a typical employer. All employees are required to have an annual appraisal of their performance, and the appraisal is done simultaneously for all employees. All employees in our sample are assessed using a common form. As in standard performance appraisals, the arrangement here is designed to capture everything relevant about an employee's job performance over the previous year. Six of the ten categories on the form refer to reasonably objective measures, such as the financial performance of individual stores against budget targets for store managers. Four of the categories refer to harder to measure, subjective issues such as the ability to work with other employees and whether the individual manifests company values. We can only observe the overall score, and that is all that is retained by the employer.

How much objective information is in a typical performance appraisal is difficult to say. Some measures, such as absenteeism, is presumably available for every job. The SMART technique for setting goals ("Specific, Measurable, Actionable, Realistic, and Targeted") appears to be used with some success by roughly 70 percent of large private sector employers (Mercer 2013), and the "measurable" attribute is about the pursuit of objective content. The Office of Personnel Management requires that appraisals for government employees include measures of "quantity, timeliness, and/or cost-effectiveness," (OPM n.d.), which points to objective measures. Supervisors are encouraged to set such goals and performance standards at the beginning of each appraisal cycle. An advantage of more objective measures is that the appraisal process becomes more transparent and credible to employees. Murphy and Cleveland (1995) assert that objective measures were common in appraisals and were often combined with subjective measures into a single score as is the case here.

The components are then aggregated by the supervisor, who has discretion as to how to weight each factor, into a single, overall measure of job performance. That score varies from 1 (lowest) to 4 (highest) and can be differentiated at the two-digit level. The instructions are explicit

[6] SEC schedule 16(b) is intended to prevent the misuse of confidential corporate information by insiders. It imposes strict liability against corporate insiders (officers, directors, and greater than 10 percent shareholders) who buy and sell or sell and buy their company's securities within a six-month period.

that the appraisal is based on the previous year's performance. During the period we observe, the company did not have a requirement that appraisal scores given by supervisors had to conform to a fixed distribution. We could not observe whether there were informal norms concerning the desired distribution.

A few other studies have made use of performance appraisal data (Medoff and Abraham 1980, 1980b; Baker et al. 1994, 1994b; Gibbs and Hendricks 2004; Castilla 2008) to examine employment outcomes where it is typically a control variable. Those studies assert as we do that the scores represent the best measures of actual performance available. The question of the validity of performance appraisal scores is difficult to assess as there is typically no measure of job performance that is more accurate than the appraisal against which its validity can be assessed. Other studies have found, however, that appraisal scores are associated with other measures of positive work outcomes such as good work attitudes (e.g., Nollen and Gaertner 1991) and that they are reliable measures.[7]

Dependent variables: We examine the relationship between performance appraisals and employment and wage outcomes. Compensation at this company begins with a base salary that is adjusted annually with a merit pay increment that then becomes part of the base salary in future years. Merit pay is designed by company policy to be based on an individual's performance as measured by the annual appraisal, although other factors such as cost of living and changes in market wages may influence merit pay outcomes as well. The merit pay measure we use is the change in the (log) of a person's salary from year to year. Annual bonus payments, which do not factor into salary, are also designed to be based on one's annual appraisal score. The overall budgets for merit pay and bonus payments each year vary based on corporate decisions. According to the company and its employee handbook, employee performance in the current year is the criterion used to determine the appraisal score at the end of that year and is also the factor used to determine merit pay and bonuses. There are no provisions for rewarding improvements in performance, although there are no rules explicitly preventing that. Nor are there any restrictions on the exact nature of the relationships between performance scores and these rewards.[8]

[7] Murphy and Cleveland (1995) discuss the issue of validity at some length and assert that the challenges to appraisal validity per se have not been persuasive. For example, showing that appraisal scores are not strongly related to objective measures of performance fails to account for the fact that objective measures in general are only one aspect of performance. They cite studies in support of appraisal validity, such as those showing that the inter-rater reliability of those observing individual performance is reasonably high.

[8] The company also has a stock option program for the employees we examine here. But option grants are identical across all employees within the same administrative level, something we control for in the analyses below. They do not vary with individual performance.

We also have data on changes in employee status: those who are promoted, demoted, dismissed, and those who quit. The first three are employer decisions. Quitting is an employee decision, albeit one that might well be influenced by appraisal outcomes: Poor performers might quit before they were fired, for example. Performance is an important factor for promotions and is the explicit factor for demotions and dismissals.

Empirical Results

Descriptive Results

TABLE 1 HERE

Are performance appraisals informative? We begin the analyses with descriptive results designed to address the general question as to whether performance appraisals are informative of individual performance specifically by examining whether they differentiate among employees. Descriptive statistics of the main variables and correlations are presented in Table 1. The density distribution in FIGURE 1 suggests that performance appraisals are informative in the important sense that they do differentiate among employees. In contrast to the common but rarely tested view noted above that the distribution of appraisal scores is sharply constrained, we see quite a different pattern here as the distribution is reasonably close to normal.

FIGURES 1 & 2 HERE

We can also reject the second version of the "do nothing" argument, that supervisors will not give poor evaluations to below average performers, by observing that the lower tail of the distribution is actually longer than that above the mean. If anything, the bias appears to be against differentiating among the higher end of the distribution.

The fact that the mean and median scores are higher than the middle of the scale might be seen as indicating that supervisors have an upward bias in their evaluations. Because the important information from these scores is the relative ranking rather than the absolute score, however, it is not clear that the position of the central tendencies relative to the scale matters or, indeed, what figure we should expect the mean and median to be. (If there is an upward bias, MacLeod (2001) concludes that such a bias in assessment and outcomes can be optimal when risk-neutral principals assess risk-averse agents subjectively.)

We also examine the argument presented above that supervisors are unwilling to give tough performance appraisal scores to employees with whom they have developed relationships by seeing what happens to appraisal scores when employees change supervisors. We cannot identify supervisor/employee pairings for most observations, but advisors in the company tell us that when employees in this sample of managers are promoted, they always change supervisors. If

social relationships inflate scores, then breaking those relationships when they are promoted and move to a new supervisor should lead to lower scores. There are other reasons as well why scores should fall following a promotion: If high scores lead to promotions, then we might expect a regression to the mean after promotion. When individuals take on a new job, we also might expect their initial performance to fall until they learn the in's and out's of the role. These factors could bias the results downward, making for a more conservative test.

Keeping those considerations in mind, we run a simple test by regressing promotions on appraisal scores after the promotion takes place. The result (not reported in the tables below) is a significant and positive relationship between prior promotions and subsequent appraisal scores in the next period after promotion. There is no evidence, therefore, that long-term relationships with the same supervisor bias appraisal scores upwards or that breaking those relationships by changing supervisors leads to lower scores. We also run a simple OLS fixed effects model examining the persistence of appraisal scores over time for the same individual (not reported here). Scores in period t account for 33 percent of the variation of scores in t+1.

Related to the topic of promotion is a common complaint that performance appraisal scores are biased toward higher-level employees. In other words, employees in management roles further up the organizational hierarchy get higher scores simply because they hold such roles. In Figure 2, we can see that it is indeed the case that average appraisal scores are higher for employees across the 10 levels in the organization (with the exception of level 3).

Whether the higher scores result from bias or from better performance is difficult to assess because we have no independent measures against which to validate the appraisal scores. But the descriptive results in Figure 3 (below) may indicate an explanation. The 10/50 ratio represents the appraisal scores in the 10^{th} percentile (bottom) as a percentage of the median score for employees at each level. The fact that it rises considerably as we move up the hierarchy may suggest that poor performers are weeded out of more important jobs. The 90/10 ratio, which is the appraisal score for the 90^{th} percentile of the distribution in that job level (top performers) to the 10^{th} percentile (bottom performers) actually falls slightly as we move up the hierarchy. The rise in average scores at higher levels of the organizational chart therefore may have to do with a reduction in low scores. We cannot rule out the hypothesis that these results stem from a bias against giving out low scores for higher level jobs, but note the earlier results that in the firm as a whole, there appears to be no reluctance to give out low scores.[9]

[9] MacLeod (2001) makes a different argument, suggesting that reduced variation in subjective scores (at the extreme to a bimodal outcome) is optimal when risk-neutral agents/subordinates are assessed by risk-neutral principals/supervisors.

In the analyses below, we examine empirically whether better performers are promoted, a result that would provide further evidence for the notion that the rise in scores as one goes up the hierarchy results from weeding out of poor performers.

Dynamics of employee performance appraisals. The final and arguably most important descriptive question is the extent to which initial human capital explains subsequent performance appraisal outcomes. If subsequent job performance is determined mainly by initial human capital or dispositions, we should expect to see little or no change in an individual's scores over time because performance should be correlated with ability at the initial hiring stage.[10] The appraisal system and management of employees in general would therefore have little effect on subsequent performance. We examine this question by observing the dynamics of performance evaluation scores, which to our knowledge have not been considered before. To capture this, we estimate a simple autoregressive model:

$$P_{it} = \alpha_i + \gamma P_{i,t-1} + \beta x_{it} + \varepsilon_{it}$$

Where P is employee i's performance evaluation at time t. $P_{i,t-1}$ is the lagged performance evaluation variable, x are control variables (including number of years in the data set [tenure], age, gender, race, and time dummies), and the term ε_{it} is the model's stochastic disturbance. Persistence in performance evaluation is capture by γ. Lower values of the adjustment parameter γ imply a smaller impact of today's evaluation on future performance appraisals - the effect declines geometrically. Suppose $\gamma = 0.2$, then this year's appraisal has only a 0.04 effect after 2 years ($Y_t=0.2Y_{t-1}= 0.2*(0.2Y_{t-2})= Y_t=0.04Y_{t-2}$. After three years the effect is negligible. The larger the parameter then the greater is the impact of today's appraisal on the future. If it is equal to one, then the best estimate of future performance appraisals is today's value; if γ is zero, then this year's appraisal is not predictive about the appraisal score in the next period.

Initially, we estimate the equation by OLS without person-fixed effects, constraining – unrealistically - all individuals to have a common intercept, $\alpha_i = \alpha$. We then estimate a simple dynamic panel data model of employee performance evaluation with person fixed-effects. We estimated the model with the GMM estimator in Arrelano-Bond (1991).[11] The fixed effects model controls for dispositional aspects of human capital, indeed, everything that an employee brought to the firm at the beginning of the period we study.

[10] Similar results could result from "Halo effects" where supervisors base subsequent appraisal scores on prior scores (see below). Note that because the data typically begins after most employees have already joined the firm, we cannot assert that the initial score we observe reflects only human capital they had before joining the firm.

[11] See also Anderson and Hsiao (1982), Bond (2002) and Blundell and Bond (1998).

The regression results are presented in Table 2. The year-to-year estimate of persistence of appraisal scores is approximately 0.51 when estimated without person-fixed effects (i.e. simple OLS). When we control for person-fixed effects (i.e. estimate the first-difference equation), however, the coefficient on the employee appraisal variable falls to 0.27: If appraisal scores were completely determined by individual dispositions (i.e., good performers were always good), the coefficient would be 1.00 and 0.00 if the scores over time were completely unrelated. The fact that scores for the same individual do vary considerably over time suggests that concerns about "halo effects" - that subsequent appraisals drive future appraisals - may be exaggerated at least in a context like this one. The idea that there are "A players" who always perform well and "C players" who performance is always mediocre is not supported by these results. They are consistent with the notion that how employees are managed matters. Our estimate suggests that steady state performance evaluation (i.e., complete 'learning' about employee ability) takes approximately four years if the magnitude of the above relationship persists over time, given the equation above.

*Do Performance Appraisals Drive Employment Outcomes**?*** The next question is to what extent do employers use appraisal outcomes as the basis for shaping employee outcomes, which is presumably the reason for conducting them in the first place. Table 3a outlines the relationship between performance appraisal scores and two important employment outcomes: merit pay increases and bonus payments in the next period. Among other things, we can examine whether the Medoff and Abraham (1980) findings that appraisal scores are not related to pay increases hold up a generation later in a context where union-based seniority systems have declined sharply and the emphasis on individual performance has increased. Bonus payments are measured separately.

Overall, we find that appraisal scores are positive and significantly related to both merit pay wage increases and bonus payments in OLS and fixed effects estimates. The fact that there are relationships between appraisal scores and these compensation outcomes does not seem surprising, but it has rarely been examined before, and it does contradict the more extreme critiques that appraisal scores fail in their basic mission of linking pay to performance.

TABLE 3a & 3b HERE

The results in Table 3a are consistent with both contractual and relational models of performance appraisals. We consider next whether supervisors reward *improvements* in

performance and punish regression across appraisal periods, which would be inconsistent with a strict settling up view of the appraisal process. Table 3b presents results showing that changes in appraisal scores between the previous year and the year when merit pay and bonuses are issued are related to those pay outcomes in the OLS and fixed effects estimates although the relationship with bonuses falls to insignificance with fixed effects.

We repeat these analyses for the important employment outcomes of promotions, demotions, dismissals, and quits in Table 4. The relationship between performance appraisal outcomes and promotions is complicated by the fact that promotions here and in most firms require that there are vacancies: Good performers cannot be promoted unless there is a position to fill. Factors other than performance in one's current job affect promotion probabilities, such as potential, matter as well. It would be very difficult for an employer in an organization like this one to guarantee that good performance in the current year would lead to promotions in that year. At best, the relationship would be probabilistic, and those relationships would be most effective if the relationship was expected to be longer-term.

In either the settling-up or the relationship view, we would expect dismissals and demotions to be associated with poorer performance. The relationship with quits should also be negative, although there are at least three possible paths for that relationship that we cannot differentiate here. Employees who are performing poorly may quit to avoid being fired (or may be allowed to quit instead of being fired), employees who intend to quit may allow their performance to decline before they do so, and factors that contribute to poor performance may also contribute to quitting, such as being at odds with an employer's norms and values. As we see in Table 1, all of these events are relatively rare, especially demotions, which might make estimation difficult. But they do occur. One percent of person-year observations experience a dismissal, for example.

TABLE 4a & 4b HERE

The results in Table 4a suggest that these outcomes are correlated to the appraisal scores: better performance is associated with a higher incidence of promotion and lower incidence of dismissals, demotions, and quits in the probit estimates. Changes in appraisal scores from the previous year in Table 4b are also significantly related to these employment outcomes in the next period. Those results are inconsistent with the notion that appraisals simply settle up performance from the previous year. Perhaps improvements in performance may be a sign of potential, driving promotions, and also indicative of a trend on the downside, encouraging employers to dismiss or demote sliding performers. Again we cannot easily identify the precise mechanism behind the relationship with quits.

Conclusions

Performance appraisals are ubiquitous in modern organizations, especially the governance structures of Anglo/US organizations (Murphy and Cleveland, 1995; Prendergast, 2002; Diaye, et al 2008; Aberdeen Group 2008;). Despite that, there is relatively little evidence on how they are used in practice and their effects on employment outcomes (Reb, 2005; Reb and Greguras 2010). We provide such evidence here. Our results indicate that performance appraisal scores are more informative than critics of appraisal systems suggest. We also find that scores vary considerably within individuals over time, suggesting at a minimum that job performance is not primarily the result of initial human capital or dispositions. Among other things, this result conflicts with arguments for dismissing employees whose appraisal scores were at the bottom of the distribution in a given year (see, e.g., Grote 2005): A low score this year may say little about one's performance next year.

More generally, appraisal scores are positively and significantly related to a range of important employment outcomes: Merit pay and bonuses, promotions, demotions and dismissals, and quits.

Perhaps most important, we find evidence that employers reward improvements in performance and that they reward different levels of performance differentially, consistent with the view that performance appraisals are not simply a means of settling up subjective aspects of prior performance. Instead, they are an adaptation to the unique, open-ended nature of employment relationships where improvements in performance matter and where employers exercise discretion in rewards that may be used to surprise subordinates.

In practice, a more realistic view of appraisals than as simply relational or simply settling-up might be a contingency notion that they vary based on the specific context of jobs. Future research that could examine appraisal practices across employers might pursue that question.

There are, of course, several caveats that apply to results like these that are based on evidence from one company. We cannot be certain that the outcomes here are representative of other employers, which means that at best our results only support the notion that performance appraisals are part of a relational notion of employment. Nor can we be certain that the jobs here are representative of typical jobs for which appraisals are used. We can say, however, that our evidence contradicts the universalistic claim of appraisals as part of a contractual process.

The increased concern about issues of endogeneity in research design is also relevant here. While fixed effects help address some of the concern, they are not a perfect solution to these problems. Future research might consider various selection issues, such as whether

employees with different levels of performance match onto different supervisors, how employees respond in the next period to scores in the current period, and so forth. Further, testing for differences between more formal contractual relationships in the workplace and more relational arrangements is not completely straight-forward. One could imagine that the different arrangements are a matter of degree rather than a bright line between them. Future models of the appraisal process as having a contingent basis, varying in their use and effects based on attributes of context, may also prove useful. Research going forward could usefully examine the structure of the performance appraisal process and how variations in arrangements affects outcomes as well as a broader range of outcomes that are influenced by it.

Overall, our analyses show that the appraisal process is more functional for the employer than the prior literature might suggest. The employment outcomes shaped by the appraisal process include some of the most fundamental issues in the economy, such as the distribution of wages, who leads organizations, and who suffers employment losses. Given that there is so little evidence about the appraisal process, we hope the evidence presented here contributes to the ongoing discussion about the role of employee performance appraisals.

Figure 1: Distribution of employee performance rating outcomes.

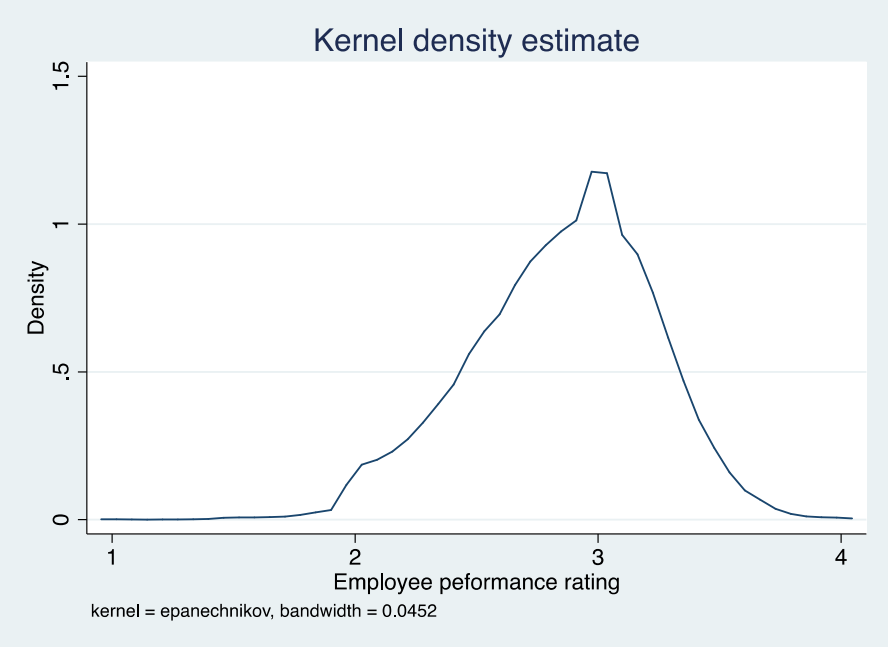

Figure 2: Employee ratings and job level

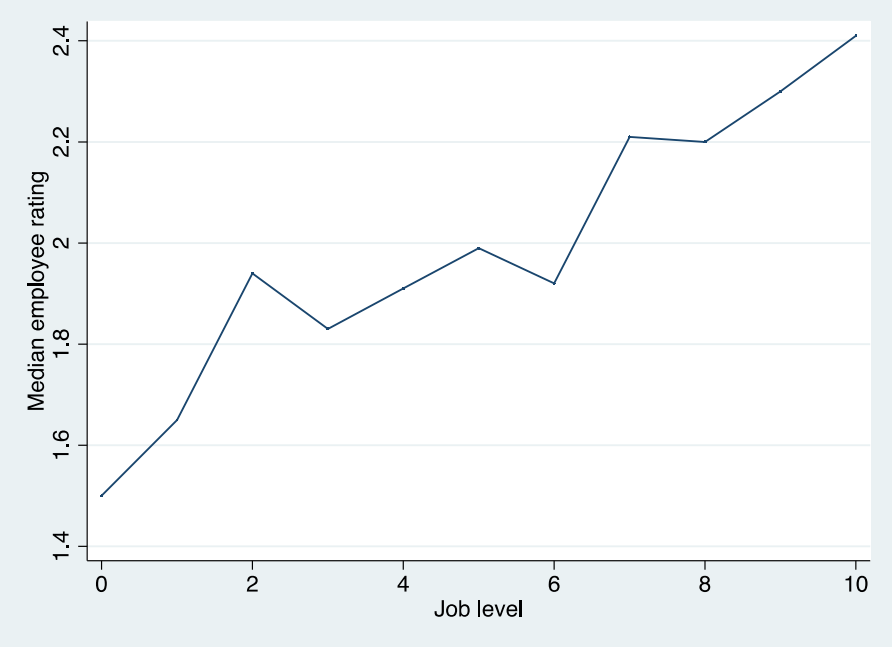

Figure 3: Employee ratings: 90th to median and 10th to median by job level

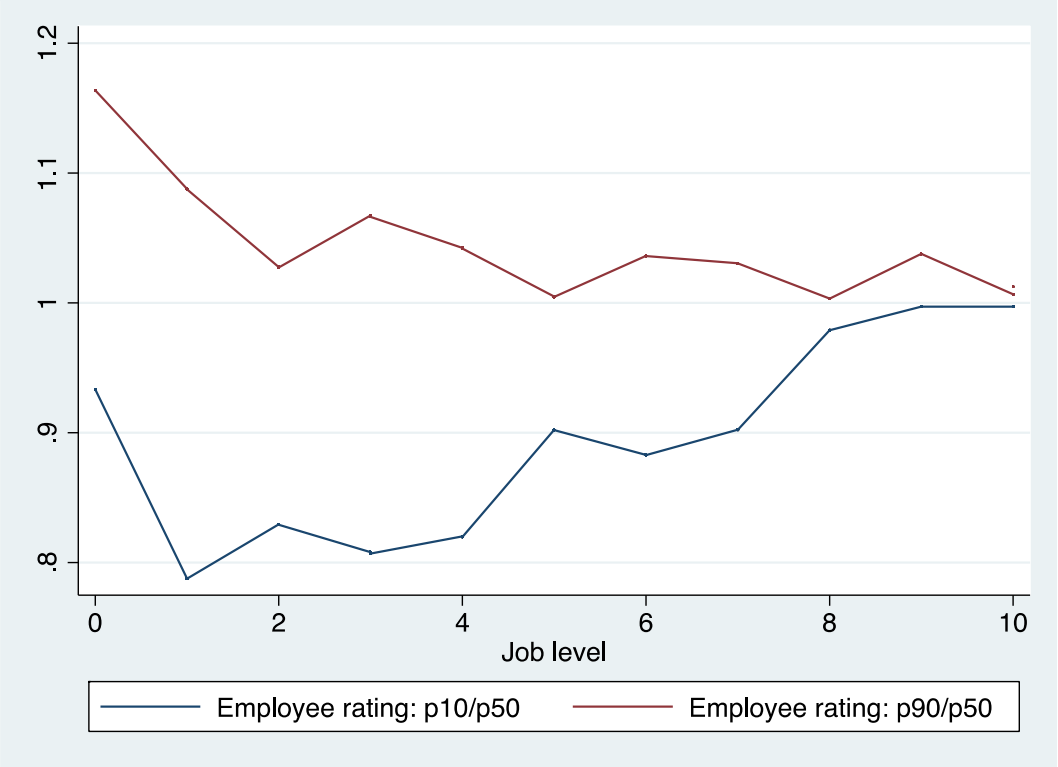

Table 1: Descriptive statistics

	Mean	S.D.	-1	-2	-3	-4	-5	-6	-7	-8	-9	-10	-11	-12
Employee rating	2.85	0.38	1											
Salary ($)	74562.84	33070.08	0.28	1										
Bonus	10601.79	15316.87	0.24	0.8	1									
Merit pay	0.05	0.08	0.17	0.06	-0.02	1								
Promotion	0.08	0.27	0.17	0.06	0	0.43	1							
Demotion	0.01	0.10	-0.04	0	0.01	-0.09	-0.03	1						
Fired	0.03	0.18	-0.14	-0.01	-0.02	-0.08	-0.04	0.01	1					
Quit	0.05	0.22	-0.06	-0.02	-0.02	-0.08	-0.03	0	-0.04	1				
White	0.90	0.30	0.06	0.04	0.05	-0.01	0.01	0.01	-0.01	-0.03	1			
Hispanic	0.03	0.16	-0.03	-0.02	-0.02	0.01	-0.01	0.00	0.00	0.02	-0.48	1		
Female	0.33	0.47	0.08	-0.1	-0.09	0.01	0.04	0.01	-0.01	0.00	0.01	-0.02	1	
Age	40.89	8.52	-0.03	0.22	0.17	-0.14	-0.08	0.01	0.06	-0.04	0.11	-0.08	-0.03	1

Table 2: The dynamics of employee performance evaluations

	OLS	Fixed Effects
Employee rating – just before appraisal	0.512***	0.277***
	(0.007)	(0.016)
Level 1	-0.026*	0.003
	(0.016)	(0.050)
Level 2	0.028*	0.137**
	(0.017)	(0.060)
Level 3	0.026	0.176***
	(0.017)	(0.065)
Level 4	0.057***	0.208***
	(0.019)	(0.065)
Level 5	0.049**	0.247***
	(0.021)	(0.071)
Level 6	0.108***	0.291***
	(0.024)	(0.076)
Level 7	0.109***	0.316***
	(0.026)	(0.082)
Level 8	0.152***	0.300***
	(0.029)	(0.089)
Level 9	0.155***	0.364***
	(0.034)	(0.099)
Level 10	0.181***	0.449***
	(0.038)	(0.114)
Log salary	0.044***	0.100
	0.878***	0.650
Constant	0.878***	0.772
	-0.154	-0.701
Year effects	Yes	Yes
Observations	18,844	14,333
M2		0.034 (0.97)

* p<0.10, ** p<0.05, *** p<0.01

Job level is a 1/0 indicator for the hierarchical level for each person in each year (column 1). Level 1 to Level 10 is an indicator variable for each hierarchical level (low to high) based on job level (job Level 0 is the excluded category). The dependent variable is individual performance evaluation in period (t). Estimation: column 1 OLS with lagged dependent variable, no person fixed effects. Column 2 contains a model with person fixed effects estimated by GMM using Arellano and Bond method. M2 is a test of second order serial correlation (Prob. > z).

Table 3a: Compensation outcomes and employee performance appraisals (levels)

	(1) Bonus: OLS	(2) Merit pay: OLS	(3) Bonus: FE	(4) Merit pay: FE
Employee rating	0.114*	0.028***	0.139**	0.033***
	(0.069)	(0.002)	(0.061)	(0.002)
Age	0.309***	-0.006***	3.709***	-0.035*
	(0.027)	(0.001)	(0.434)	(0.021)
Age squared	-0.004***	0.000***	-0.007***	0.000***
	(0.000)	(0.000)	(0.001)	(0.000)
Job level	1.120***	0.005***	0.501***	0.045***
	(0.015)	(0.000)	(0.046)	(0.003)
Constant	-2.145***	0.086***	-129.252***	0.989
	(0.598)	(0.016)	(16.732)	(0.806)
Control variables	Yes	Yes	No	No
Observations	18,844	18,842	18,844	18,842
R-squared	0.314	0.057	0.116	0.102
Number of persons			4,408	4,408

* p<0.10, ** p<0.05, *** p<0.01

OLS estimates presented with heterosedastic robust t-statistics in parenthesis; FE are person-specific fixed effects estimates with heterosedastic robust t-statistics in parenthesis. Employee rating is in the range 1 to 4. Control variables include White, Hispanic, and Female are indicator variables. Age is in years. Job level is an integer associated with the hierarchical level assigned to the employee in the given year. Regressions contain a set of time dummies and estimated over the years 2001 to 2007.

Table 3b: Pay outcomes and employee performance appraisals (percentage change)

	(1) Bonus: OLS	(2) Merit pay: OLS	(3) Bonus: FE	(4) Merit pay: FE
Employee rating	0.492***	0.052***	-0.062	0.049***
(% change)	(0.174)	(0.005)	(0.094)	(0.005)
Age	0.308***	-0.006***	3.712***	-0.033
	(0.027)	(0.001)	(0.435)	(0.021)
Age squared	-0.004***	0.000***	-0.007***	0.000***
	(0.000)	(0.000)	(0.001)	(0.000)
Job level	1.127***	0.007***	0.506***	0.046***
	(0.014)	(0.000)	(0.046)	(0.003)
Control variables	Yes	Yes	No	No
Observations	18,844	18,842	18,844	18,842
R-squared	0.314	0.049	0.115	0.097
Number of persons			4,408	4,408

* p<0.10, ** p<0.05, *** p<0.01

OLS estimates presented with heterosedastic robust t-statistics in parenthesis; FE are person-specific fixed effects estimates with heteroskedastic robust t-statistics in parenthesis. Employee rating is in the range 1 to 4. Control variables include White, Hispanic, and Female indicator variables. Age is in years. Job level is an integer associated with the hierarchical level assigned to the employee in the given year. Regressions contain a set of time dummies and estimated over the years 2001 to 2007. The Employee rating percentage change is calculated as: [R(t)-R(t-1)/R(t-1)] where R(t) is employee rating at time t and so on.

Table 4a: Labor market outcomes and employee performance appraisals

	(1) Promotion	(2) Demotion	(3) Dismissal	(4) Quits
Employee rating (t)	0.137***	-0.011***	-0.049***	-0.032***
	(0.005)	(0.002)	(0.002)	(0.003)
Age	-0.009***	0.000	0.002**	-0.008***
	(0.002)	(0.001)	(0.001)	(0.001)
Age squared	0.000***	-0.000	-0.000	0.000***
	(0.000)	(0.000)	(0.000)	(0.000)
Control variables	Yes	Yes	Yes	Yes
Observations	20,741	20,741	23,412	23,412

Models include a constant and time dummies; Control variables include White, Hispanic, and Female indicator variables. These are not reported in the tables. Marginal effects from a probit model estimation are reported. Other variables defined in the text.

Table 4b: Labor market outcomes and employee performance appraisals

	(1) Promotion	(2) Demotion	(3) Dismissal	(4) Quits
Employee rating (% change)	0.134***	-0.031***	-0.091***	-0.067***
	(0.014)	(0.007)	(0.011)	(0.012)
Age	-0.014***	-0.000	0.001	-0.012***
	(0.002)	(0.001)	(0.001)	(0.001)
Age squared	0.000***	0.000	0.000	0.000***
	(0.000)	(0.000)	(0.000)	(0.000)
Demographic variables	Yes	Yes	Yes	Yes
Observations	18,844	18,844	18,844	18,844

Models include a constant and time dummies not reported. Marginal effects from a probit model estimation are reported. Rating percentage change is [R(t)-R(t-1)/R(t-1)] where R(t) is employee rating at time t and so on. Variables are defined in the text. Control variables include White, Hispanic, and Female indicator variables.

References

Aberdeen Group. 2010. *"Employee Performance Management: The Alpha and the Omega of Talent Strategy and Business Execution." Boston*, MA. http://www.aberdeen.com/aberdeen-library/6361/RA-employee-performance-management.aspx.

Addison, John T., and Clive R. Belfield. 2008. "The Determinants of Performance Appraisal Systems: A Note (do Brown and Heywood's Results for Australia Hold up for Britain?)." *British Journal of Industrial Relations* 46 (3): 521–31.

Almeda, David. 2009. "Investment or Identification? The Decision to Exercise Stock Options as an Indicator of Organizational Identification" PhD Graduate School of Education, University of Pennsylvania.

Anderson, Theodore Wilbur, and Cheng Hsiao. 1982. "Formulation and Estimation of Dynamic Models Using Panel Data." *Journal of Econometrics* 18 (1): 47–82.

Aral, Sinan, Erik Brynjolfsson, and Lynn Wu. 2012. "Three-Way Complementarities: Performance Pay, Human Resource Analytics, and Information Technology." *Management Science* 58 (5): 913–31.

Arellano, Manuel, and Stephen Bond. 1991. "Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations." *The Review of Economic Studies* 58 (2): 277–97.

Arya, Anil, and Brian Mittendorf. 2011. "The Benefits of Aggregate Performance Metrics in the Presence of Career Concerns." *Management Science* 57 (8): 1424–37.

Atleson, James B. 1983. *Values and Assumptions in American Labor Law*. Amherst, Mass.: University of Massachusetts Press.

Baker, George, Michael Gibbs, and Bengt Holmstrom. 1994. "The Wage Policy of a Firm." *The Quarterly Journal of Economics* 109 (4): 921–55. doi:10.2307/2118352.

Baker, George, Robert Gibbons, and Kevin J. Murphy. 1994. "Subjective Performance Measures in Optimal Incentive Contracts." *The Quarterly Journal of Economics* 109 (4): 1125–56. doi:10.2307/2118358.

Baker, George P. 1992. "Incentive Contracts and Performance Measurement." *Journal of Political Economy* 100 (3): 598–614. doi:10.1086/261831.

Bartel, Ann P. 1995. "Training, Wage Growth, and Job Performance: Evidence from a Company Database." *Journal of Labor Economics* 13 (3): 401–25. doi:10.1086/298380.

John Bernardin, H., Stephanie Thomason, M. Ronald Buckley, and Jeffrey S. Kane. 2015. "Rater Rating-Level Bias and Accuracy in Performance Appraisals: The Impact OF Rater Personality, Performance Management Competence, and Rater Accountability." *Human Resource Management* 55(2): 321-340

Bloom, Nicholas, and John Van Reenen. 2010. "Why Do Management Practices Differ across Firms and Countries?" *The Journal of Economic Perspectives* 24 (1): 203–24.

Blundell, Richard, and Stephen Bond. 1998. "Initial Conditions and Moment Restrictions in Dynamic Panel Data Models." *Journal of Econometrics* 87 (1): 115–43.

Bond, Stephen. (2002) 'Dynamic panel data models: a guide to micro data methods and practice', Portugese Economic Journal 1, 141-162.

Borjas, George. 2010. Labor Economics. New York: McGraw-Hill.

Brickley, James A., Clifford W. Smith, Jerold L. Zimmerman, Zhiqiang Zhang, and Chunxiang Wang. 2001. *Managerial Economics and Organizational Architecture*. Vol. 4. McGraw-Hill/Irwin. http://works.bepress.com/clifford_smith/.

Brown, Michelle, and John S. Heywood. 2005. "Performance Appraisal Systems: Determinants and Change." *British Journal of Industrial Relations* 43 (4): 659–79. doi:10.1111/j.1467-8543.2005.00478.x.

Brodie, Douglas. 2011. "How Relational Is the Employment Contract?" *Industrial Law Journal* 40 (3): 232–53. doi:10.1093/indlaw/dwr009.

Brown, John Howard. 2011. "Contracts Versus Price Discrimination: Evidence From the SONJ Case." *Review of Industrial Organization* 38 (3): 235–43. doi:10.1007/s11151-011-9285-9.

Buehler, Stefan, and Dennis L Gärtner. 2013. "Making Sense of Nonbinding Retail-Price Recommendations." *American Economic Review* 103 (1): 335–59. doi:10.1257/aer.103.1.335.

Bull, Clive. 1987. "The Existence of Self-Enforcing Implicit Contracts." *The Quarterly Journal of Economics*, 102(1):147–59.

[BLS] Bureau of Labor Statistics. 2014. "Employed Persons by Occupation, Sex, and Age." Accessed at http://www.bls.gov/cps/cpsaat09.htm.

Castilla, Emilio J. 2008. "Gender, Race, and Meritocracy in Organizational Careers." *American Journal of Sociology* 113 (6): 1479–1526. doi:10.1086/588738.

Cheung, Sai On, Kenneth T. Yiu, and Pui Shan Chim. 2006. "How Relational Are Construction Contracts?" *Journal of Professional Issues in Engineering Education and Practice* 132 (1): 48–56.

Culbert, Samual A. 2008. "Get Rid of the Performance Review! It Destroys Morale, Kills Teamwork and Hurts the Bottom Line; And That's Just for Starters." *Wall Street Journal at: Http://online. Wsj. com/article/SB1224263188* 74844933. http://www.performance-appraisals.org/cgi-bin/links/jump.cgi?ID=10666

Davis, Tom, and Michael Landa. 1999. "A Contrary Look at Employee Performance Appraisal." *Canadian Manager*, September. https://www.highbeam.com/doc/1G1-58436540.html.

Davis, Tom, and Michael Landa .1999. "A Contrary Look at Employee Performance Appraisal." *Canadian Manager* 24: 18–19.

Diaye, Marc-Anthony, N. Greenan, and M. Urdanivia (2008). "Subjective Evaluation of Performance through Individual Evaluation Interview: Empirical Evidence from France". In: S. Bender, J. Lane, K. Shaw, F. Andersson, and T. von Wachter (eds.). *The Analysis of Firms and Employees: Quantitative and Qualitative Approaches.* NBER and University of Chicago Press, pp. 107-131.

Piore, Michael J., and Peter Doeringer. 1971. "Internal Labor Markets and Manpower Adjustment." *New York: DC Heath and Company*.

Dohmen, Thomas J. 2004. "Performance, Seniority, and Wages: Formal Salary Systems and Individual Earnings Profiles." *Labour Economics* 11 (6): 741–63.

Feltham, Gerald A., and Jim Xie. 1994. "Performance Measure Congruity and Diversity in Multi-Task Principal/agent Relations." *Accounting Review*, 69: 429–53.

Fenton-O'Creevy, Mark, Nigel Nicholson, Emma Soane, and Paul Willman. 2004. *Traders: Risks, Decisions and Management in Financial Markets*. Oxford University Press.

Foss, Nicolai J., and Keld Laursen. 2005. "Performance Pay, Delegation and Multitasking under Uncertainty and Innovativeness: An Empirical Investigation." *Journal of Economic Behavior & Organization* 58 (2): 246–76.

Gibbs, Michael, and Wallace Hendricks. 2004. "Do Formal Salary Systems Really Matter?" *Industrial & Labor Relations Review* 58 (1): 71–93.

Gibbs, Michael, Kenneth A. Merchant, Wim A. Van der Stede, and Mark E. Vargus. 2004. "Determinants and Effects of Subjectivity in Incentives." *The Accounting Review* 79 (2): 409–36.

Green, Les. 2003. Legal Positivism. Stanford Encyclopedia of Philosophy. http://plato.stanford.edu/entries/legal-positivism/

Grote, Richard C. 2005. *Forced Ranking: Making Performance Management Work*. Harvard Business Press.

Grund, Christian, and Judith Przemeck. 2012. "Subjective Performance Appraisal and Inequality Aversion." *Applied Economics* 44 (17): 2149–55.

Harari, Michael B., Cort W. Rudolph, and Andrew J. Laginess. 2015. "Does Rater Personality Matter? A Meta-Analysis of Rater Big Five–performance Rating Relationships." *Journal of Occupational and Organizational Psychology* 88 (2): 387–414.

Hart, Oliver. 1995. *Firms, Contracts, and Financial Structure*. Clarendon Press.

Holmstrom, Bengt, and Paul Milgrom. 1991. "Multitask Principal-Agent Analyses: Incentive Contracts, Asset Ownership, and Job Design." *Journal of Law, Economics, & Organization* 7: 24–52.

Hölmstrom, Bengt. 1979. "Moral Hazard and Observability." *The Bell Journal of Economics*, 74–91.

Holmström, Bengt. 1999. "Managerial Incentive Problems: A Dynamic Perspective." *The Review of Economic Studies* 66 (1): 169–82.

Ichniowski, Casey, Kathryn Shaw, and Giovanna Prennushi. 1997. "The Effects of Human Resource Management Practices on Productivity: A Study of Steel Finishing Lines." *The American Economic Review*, 291–313.

Ilgen, Daniel R., Janet L. Barnes-Farrell, and David B. McKellin. 1993. "Performance Appraisal Process Research in the 1980s: What Has It Contributed to Appraisals in Use?" *Organizational Behavior and Human Decision Processes* 54 (3): 321–68.

Indjejikian, Raffi J. 1999. "Performance Evaluation and Compensation Research: An Agency Perspective." *Accounting Horizons* 13 (2): 147–57.

Indjejikian, Raffi J., and Michal Matějka. 2011. "Accounting Decentralization and Performance Evaluation of Business Unit Managers." *The Accounting Review* 87 (1): 261–90.

Kahn, Lawrence M., and Peter D. Sherer. 1990. "Contingent Pay and Managerial Performance." *Industrial & Labor Relations Review* 43 (3): 107S – 120S.

Klein, Benjamin, Robert G. Crawford, and Armen A. Alchian. 1978. "Vertical Integration, Appropriable Rents, and the Competitive Contracting Process." *The Journal of Law & Economics* 21 (2): 297–326.

Lazear, Edward P. 1986. "Salaries and Piece Rates." *Journal of Business*, 405–31.

Lazear, Edward P. 1990. "The Timing of Raises and Other Payments." In *Carnegie-Rochester Conference Series on Public Policy*, 33:13–48. Elsevier. http://www.sciencedirect.com/science/article/pii/0167223190900056.

Lazear, Edward P. 2000. "Performance Pay and Productivity." *American Economic Review* 90 (5): 1346–61.

Levin, Jonathan. 2003. "Relational Incentive Contracts." *The American Economic Review* 93 (3): 835–57.

Likert, Rensis. 1959. *Motivational Approach to Management Development*. Harvard Business Review. July-August p. 75-88.

Longenecker, Clinton O., Henry P. Sims Jr, and Dennis A. Gioia. 1987. "Behind the Mask: The Politics of Employee Appraisal." *The Academy of Management Executive (1987-1989)*, 1(3): 183–93.

Longenecker, Clinton O., Laurence S. Fink, and Sheri Caldwell. 2014. "Current US Trends in Formal Performance Appraisal: Practices and opportunities–Part II." *Industrial and Commercial Training* 46 (7): 393–99.

Bentley MacLeod, W. 2003. "Optimal Contracting with Subjective Evaluation." *The American Economic Review* 93 (1): 216–40.

MacLeod, B. and D. Parent (1999). Job Characteristics and the Form of Compensation. Research in Labor Economics 18: 177-242.

Diaye, Marc-Arthur, Nathalie Greenan, and Michal W. Urdanivia. 2008. "Subjective Evaluation of Performance and Evaluation Interview: Empirical Evidence from France." In *The Analysis of Firms and Employees: Quantitative and Qualitative Approaches*, 107–31. University of Chicago Press. http://www.nber.org/chapters/c9113.pdf.

Macaulay, Stewart. 1963. "Non-Contractual Relations in Business: A Preliminary Study." *American Sociological Review*, 28(1): 55–67.

Medoff, James L., and Katharine G. Abraham. 1980. "Experience, Performance, and Earnings". *Quarterly Journal of Economics*, 95(4): 703-736.

Medoff, James L., and Katharine G. Abraham. 1981. "Are Those Paid More Really More Productive? The Case of Experience." *Journal of Human Resources*, 16: 186–216.

Mercer. 2013. "2013 Global Performance Management Survey Report". http://www.mercer.com/content/dam/mercer/attachments/global/Talent/Assess-BrochurePerfMgmt.pdf

Mero, Neal P., Stephan J. Motowidlo, and Alexandra L. Anna. 2003. "Effects of Accountability on Rating Behavior and Rater Accuracy." *Journal of Applied Social Psychology* 33 (12): 2493–2514.

Michaels, Ed, Helen Handfield-Jones, and Beth Axelrod. 2001. *The War for Talent*. Boston: Harvard Business Press.

Milgrom, Paul R. 1988. "Employment Contracts, Influence Activities, and Efficient Organization Design." *The Journal of Political Economy*, 96: 42–60.

Milgrom, Paul, and John Roberts. 1988. "An Economic Approach to Influence Activities in Organizations." *American Journal of Sociology*, S154–79.

Milkovich, George T, Alexandra K Wigdor, National Research Council (U.S.), and Committee on Performance Appraisal for Merit Pay. 1991. *Pay for Performance: Evaluating Performance Appraisal and Merit Pay*. Washington, D.C.: National Academy Press.

Mosley, Donald C., Leon C. Megginson, and Paul H. Pietri. 1985. *Supervisory Management: The Art of Working with and through People*. Thomson South-Western.

Murphy, Kevin J., and Paul Oyer. 2001. "Discretion in Executive Incentive Contracts: Theory and Evidence." *Available at SSRN 294829*. http://papers.ssrn.com/sol3/papers.cfm?abstract_id=294829.

Murphy, Kevin R., and Jeanette Cleveland. 1995. *Understanding Performance Appraisal: Social, Organizational, and Goal-Based Perspectives*. Oaks, CA: Sage Publications.

Nagar, Venky. 2002. "Delegation and Incentive Compensation." *The Accounting Review* 77 (2): 379–95.

Nollen, Stanley D., and Karen N. Gaertner. 1991. "Effects of Skill and Attitudes on Employee Performance and Earnings." *Industrial Relations: A Journal of Economy and Society* 30 (3): 435–55.

[OPM] Office of Personnel Management (n.d.) Performance Appraisal Assessment Tool. https://www.opm.gov/policy-data-oversight/performance-management/measuring/gs-paat-instructions.pdf

Ortega, Jaime. 2009. "Employee Discretion and Performance Pay." *The Accounting Review* 84 (2): 589–612.

Peretz, Hilla, and Yitzhak Fried. 2012. "National Cultures, Performance Appraisal Practices, and Organizational Absenteeism and Turnover: A Study across 21 Countries." *Journal of Applied Psychology* 97 (2): 448-459.

Pichler, Shaun. 2012. "The Social Context of Performance Appraisal and Appraisal Reactions: A Meta-Analysis." *Human Resource Management* 51 (5): 709–32.

Prendergast, Canice. 1993. "A Theory of 'Yes Men.'" *The American Economic Review*, 757–70.

Prendergast, Canice. 1999. "The Provision of Incentives in Firms." *Journal of Economic Literature* 37 (1): 7–63.

Prendergast, Canice, and Robert H. Topel. 1996. "Favoritism in Organizations." *Journal of Political Economy* 104 (5): 958–78. doi:10.1086/262048.

Prendergast, Canice. 1999. "The Provision of Incentives in Firms." *Journal of Economic Literature* 37 (1): 7–63.

Prendergast, Canice. 2002. "Uncertainty and Incentives." *Journal of Labor Economics* 20 (S2): S115–37.

Rayo, Luis. 2007. "Relational Incentives and Moral Hazard in Teams." *The Review of Economic Studies* 74 (3): 937–63.

Reb, Jochen, and Gary J. Greguras. 2010. "Understanding Performance Ratings: Dynamic Performance, Attributions, and Rating Purpose." *Journal of Applied Psychology* 95 (1): 213-220.

Reb, Jochen. 2005. "Evaluating Dynamic Performance: Effects of Trend, Mean, and Variation of *Performance on Performance Ratings". Academy of Management Proceedings, pJ1-J6,*

Schwab, Donald P., and Craig A. Olson. 1990. "Merit Pay Practices: Implications for Pay-Performance Relationships." *Industrial & Labor Relations Review* 43 (3): 237S – 255S.

Selden, Sally Coleman, Patricia Wallace Ingraham, and Willow Jacobson. 2001. "Human Resource Practices in State Government: Findings from a National Survey." *Public Administration Review* 61 (5): 598–607.

Seltzer, Andrew, and David Merrett. 2000. "Personnel Policies at the Union Bank of Australia: Evidence from the 1888–1900 Entry Cohorts." *Journal of Labor Economics* 18 (4): 573–613.

Simon, Herbert A. 1951. "A Formal Theory of the Employment Relationship." *Econometrica: Journal of the Econometric Society*, 19(3): 293–305.

Thompson, Frank J., and Beryl A. Radin. 1997. *Reinventing Public Personnel Management: The Winter and Gore Initiatives*. In Public Personnel Management: Current Concerns, Future Challenges, edited by Carolyn Ban and Norma M. Riccucci. New York: Longman Publishers.

Williamson, Oliver E. 1985. *The Economic Institutions of Capitalism Firms Markets Relational Contracting*. Free Press.

Yang, Huanxing. 2008. "Efficiency Wages and Subjective Performance Pay." *Economic Inquiry* 46 (2): 179–96.