The Implications of Rating Systems on Workforce Performance

Christopher Green    Morvarid Rahmani

Scheller College of Business, Georgia Institute of Technology, Atlanta, GA 30308
christopher.green@scheller.gatech.edu; morvarid.rahmani@scheller.gatech.edu

Accepted for publication in IISE Transactions

Abstract

Enhancing workforce performance is the key to success for professional firms. Firms often evaluate workers based on their performance compared to their peers or against an objective standard. Which of these rating systems leads to higher workforce performance? To answer this question, we construct game-theoretic models of two performance rating systems: (i) a Relative rating system where workers compete with each other for a constrained number of high ratings, and (ii) an Absolute rating system where workers are awarded high ratings by performing at or above a standard threshold. We derive the workers’ equilibrium performance as a function of their ability and the characteristics of the rating pool. From a firm’s perspective, we find that an Absolute rating system can lead to higher performance than a Relative rating system when the rating pool size is small or the workers’ cost of effort relative to their efficiency rate is low, and the reverse holds true otherwise. When considering the workers’ perspective, we find that higher ability workers prefer an Absolute system due to its predictable nature, while lower ability workers prefer a Relative system as it provides them an opportunity to outperform other workers.

Keywords: workforce management; rating system, performance evaluation; game theory

1. Introduction

Firms need to continually enhance their workforce performance to maintain a competitive edge in a tight labor market. Improving workers’ performance is challenging in professional firms because workers’ abilities are not fully known and efforts are not observable. Thus, firms often institute internal promotion systems as a lever to motivate professional workers to improve their performance (Rohman et al., 2018). That is, they adopt a rating system that allows them to differentiate between high and low performing workers (Martin and Schmidt, 2010; Bidwell, 2011; Keller and Meaney,
2017) with the intention to promote high performing workers to more advanced positions. Rating systems either compare workers against one another, which are referred to as Relative, or against a standard, which are referred to as Absolute (Cascio and Aguinis, 2018, Ch. 5). Within these two general types, there can be several methods of evaluation.¹

Relative rating systems are competitive and constrain the firm to award high ratings to a certain proportion of workers. For instance, the United States Army currently uses a Relative rating system where only the top 49% of commissioned officers can receive a high rating referred to as a “Most Qualified” report. These ratings are the most discriminating factor in the Army’s centralized selection process which promotes individuals into some of the highest ranks of the United States Government (Department of the Army, 2019). Although some firms have moved away from this type of system due to legal actions and unintended impacts on the workplace, many firms have continued with this practice because they believe that it is more effective in boosting workers’ performance (Sloan et al., 2017).

Absolute rating systems are non-competitive and evaluate workers’ performance against an objective standard. For instance, Intel and Google’s Objectives and Key Results (OKR) or Adobe’s “Check-in” system requires supervisors and workers to discuss workers’ performance against objective standards to help them attain the skills they need to continue growing and improving at the company (Meinert, 2015; Doerr, 2018). We consider an Absolute rating system with a single objective standard, where workers who perform to or exceed a performance threshold are given a high rating. Many experts and workers believe that Absolute rating systems are more fair, avoid the effects of bias, and prevent sabotage in the workplace (Roch et al., 2007). However, this rating system provides less control over the number (or proportion) of workers that receive high ratings, and it could also be less useful in differentiating the highest performers.

Researchers in economics, operations management, and organizational behavior have studied the design and performance of various rating systems. While the organizational behavior literature has mostly studied the psychological traits and inclination of a rater and/or ratee for a specific rating system, the economics and operations management literatures have mostly studied compensation

¹For instance, rating systems such as behavioral checklists and graphic rating scales share similar characteristics as with the Absolute system in the sense that workers are evaluated against an objective standard. However, rating systems such as rank ordering, paired comparison, and forced distribution share similar characteristics as with the Relative system in the sense that a worker’s performance is compared against that of his peers (Cascio and Aguinis, 2018, Ch. 5).
schemes under different rating systems (see detailed discussions in §2). However, in the case of internal promotion where workers’ incentives could be non-monetary and immutable, there is a lack of normative studies comparing the direct impact of rating systems on the workforce performance. This paper fills this gap by addressing the following research questions: (i) From the firm’s perspective, what rating system (Relative or Absolute) leads to a higher workforce performance? (ii) How does the optimality of a rating system depend on characteristics of rating pools and workers? To gain a comprehensive understanding of these systems, we also ask: (iii) What is the preferred rating system from the workers’ perspective?

We develop game-theoretic models of Relative and Absolute rating systems. In both systems, a worker’s performance is a function of his ability and choice of costly effort. Workers know their own ability and the overall distribution of other workers’ abilities, but they cannot observe each other’s effort. Similarly, the firm cannot observe the workers’ abilities and efforts, but it can verify the overall performance of each worker at the end of the rating period. Under a Relative system, only a fraction of workers with the highest performance receive high ratings; whereas under an Absolute system, those workers who perform at or above a threshold will receive high ratings. The firm prefers a rating system that leads to a higher overall performance of the rating pool, while workers prefer a rating system that leads to a higher individual payoff for them. We characterize the workers’ equilibrium choices and compare the performance of the two systems.

We find that workers with mid-level abilities exert the highest effort in both systems, but the magnitude of their efforts can be higher in an Absolute system than in a Relative system. Accordingly, an Absolute system can lead to higher overall performance when the rating pool size is small or the workers’ cost of effort relative to their efficiency rate is low, as in many routine jobs. In contrast, a Relative system can lead to better performance when the rating pool size is large or the workers’ cost of effort relative to their efficiency rate is high, as in many knowledge-intensive jobs. Considering the workers’ perspective, we find that workers with high abilities prefer an Absolute system because of the predictability provided by the publicly declared threshold. In contrast, workers with lower abilities find a Relative system beneficial because although they may not be able to perform to the Absolute threshold, there is a chance they may outperform their peers in a Relative system. Not only do the low ability workers prefer a Relative system, their performance is also higher in a Relative system than in an Absolute system, indicating an alignment between the firm’s
and workers’ perspectives. In contrast, high ability workers, who prefer an Absolute system, could have higher performance in a Relative system, indicating a misalignment between the firm’s and workers’ perspectives. We find that firms can improve the performance of rating systems and also enhance the alignment between the firm’s and workers’ perspective by optimally setting thresholds and/or rewards in a way that exploits the difference between the two systems. Specifically, under a Relative system, the firm can benefit from setting higher rewards and offering a fewer number of high ratings to motivate high ability workers to exert high efforts. In contrast, under an Absolute system, the firm can benefit from setting lower rewards and offering a larger number of high ratings to prompt low ability workers to exert efforts.

The remainder of the paper is organized as follows. We review the related literature in §2 and present models of Relative and Absolute rating systems in §3. Our main results are presented in §4. We provide an extensive numerical study based on the context of the U.S. Army in §5. We conclude with a summary of managerial insights and directions for future research in §6. All proofs and technical details are presented in the appendix.

2. Literature Review

In this paper, we contribute to the economic, operations management, and organizational behavior literatures that study performance rating systems.

Economic and Operations Management Models of Rating Systems: In professional work environments, abilities are not fully known (Kwon and Yoo, 2017) and efforts are not observable (Holmstrom et al., 1979), which result in inefficiencies due to moral hazard. There exists a large strand of literature on designing compensation schemes to improve workforce performance and mitigate moral hazard (e.g., Wu et al. 2014; Zhang 2016; Crama et al. 2019; Rahmani and Ramachandran 2020). In addition to (or instead of) compensation schemes, firms can employ other organizational levers such as promotion (administered via rating systems) to improve workforce performance (Lazear and Rosen, 1981; Waldman, 2013; Barlevy and Neal, 2019). Research on performance rating systems can be categorized into three groups, depending on their focus on competitive or non-competitive rating systems.
A group of studies have focused on only competitive rating systems. For instance, Kwon (2013) and Miklós-Thal and Ullrich (2014) show that when workers compete for promotions, their efforts can increase with manager’s belief precision of their abilities. In addition, through simulation, Scullen et al. (2005) show that a relative system with forced distribution performs better when the percentage of workers that can get a low rating is small and that voluntary turnover is low. Similarly, Evans (2018) show that a relative system can be more accurate and effective when the size of the rating pool is large. In contrast, Harbring and Irlenbusch (2003) show that performance tends to increase in a relative system when a proportion of workers that can get high ratings is high. Empirically, Casas-Arce and Martinez-Jerez (2009) showed that in relative performance tournaments, higher ability workers decrease effort as the probability of winning increases and in general, workers decrease effort as the number of participants decreases (without an increase in prizes). We contribute to this stream of research by comparing the performance of a competitive rating system to that of a non-competitive rating system.

Another group of studies have focused on only non-competitive rating systems. For instance, Ghosh and Waldman (2010) show that standard promotion (where there is no deadline for promotion) perform better than up-or-out promotions when the firm-specific human capital is high. Corgnet et al. (2015) and Corgnet et al. (2018) show that setting a standard goal can motivate workers to exert higher effort beyond what is achieved by using solely monetary incentives. Considering a team setting, Fan and Gómez-Miñambres (2020) show that setting standard goals can increase team performance, especially when goals are challenging but attainable for weak-link workers. In a meta-analysis, Rodgers and Hunter (1991) find that a non-competitive rating system in the form of Management by Objective can improve performance between 6% and 56% depending on the commitment level of managers. We contribute to this stream of research by generating insights on the performance of a non-competitive rating system as compared to a competitive rating system.

Our work is closer to a third group of studies that compares different rating systems (e.g., Lazear and Rosen, 1981; Green and Stokey, 1983). In the seminal work on tournaments as labor contracts, Lazear and Rosen (1981) show that, under certain conditions, compensating workers on the basis of their relative rank can yield similar performance as that generated by efficient piece-rate compensation in a pool of two workers. Green and Stokey (1983) extend the work of Lazear and Rosen (1981) for any number of workers and show that the relative compensation
scheme can dominate the independent contract when the workers’ outputs are stochastic; otherwise, independent contracts can result in better outcomes. Nalebuff and Stiglitz (1983) show that a competitive compensation scheme performs better than a piece-rate compensation especially when task uncertainty is high. Most recently, Jain et al. (2019) compare outcome- and ranking-based compensation schemes for a pool of two workers with observable efforts. They show that a ranking-based system can perform better when the task does not require a high level of teamwork. Otherwise, an outcome-based system performs better. Our work differs from these studies in three main ways: First, while the above papers have studied compensation schemes under different rating systems, we capture situations where incentives could be non-monetary and immutable (which is synonymous with internal promotions) and analyze the direct effects of rating systems on workforce performance. Second, unlike the above papers that considered a pool of homogeneous workers, we consider heterogeneous workers (two or greater) who possess private information about their ability to accomplish their job. Finally, we compare the rating systems from both the firm’s perspective (performance) and the workers’ perspective (payoff) and identify areas of congruence between these perspectives.

This paper is also related to an adjacent stream of research that studies competition among agents in contests. These models share similarities with the Relative rating system, as participants are competing against each other and only a fraction of them can receive rewards. This stream of research has generally focused on the design of contests in terms of size of the reward(s) (Bimpikis et al., 2019; Moldovanu and Sela, 2006), participation fee (Terwiesch and Xu, 2008; Körpeoglu and Cho, 2017), and feedback mechanism (Mihm and Schlapp, 2018). We contribute to this stream of literature by comparing the performance of the Relative and Absolute rating systems, a question that is not answered by the above papers. In addition, we compare the performance of these rating systems with comparative thresholds (that result in the same expected number of high ratings in both systems) and optimal thresholds under each system, and generate insights on how firms can use these thresholds as additional levers to improve workforce performance.

**Organizational Behavior and Psychology of Rating Systems:** Performance of different rating systems have also been widely studied in organizational behavior and psychology literature. Table 1 summarizes a sample of studies that considered competitive and/or non-competitive rating
Table 1: Rating Systems in Organizational Behavior Literature

<table>
<thead>
<tr>
<th>Reference</th>
<th>Competitive</th>
<th>Non-competitive</th>
</tr>
</thead>
<tbody>
<tr>
<td>Landy and Farr (1980)</td>
<td>Derived</td>
<td>Direct</td>
</tr>
<tr>
<td>Siegel (1982)</td>
<td>Paired Comparison</td>
<td></td>
</tr>
<tr>
<td>Rodgers and Hunter (1991)</td>
<td></td>
<td>Management by Objective</td>
</tr>
<tr>
<td>Goffin et al. (1996, 2009); Cascio and Aguinis (2018); Roch et al. (2007)</td>
<td>Relative</td>
<td>Absolute</td>
</tr>
<tr>
<td>Wagner and Goffin (1997)</td>
<td>Comparative</td>
<td>Absolute</td>
</tr>
<tr>
<td>Locke and Latham (2002); Fan and Gómez-Miñambres (2020)</td>
<td></td>
<td>Goal Setting</td>
</tr>
<tr>
<td>Scullen et al. (2005); Berger et al. (2013); Blume et al. (2009, 2013)</td>
<td>Forced Distribution</td>
<td></td>
</tr>
<tr>
<td>Orrison et al. (2004)</td>
<td>Promotion Tournament</td>
<td></td>
</tr>
<tr>
<td>Harbring and Lünser (2008); Gill et al. (2019)</td>
<td>Rank Order Tournament</td>
<td></td>
</tr>
<tr>
<td>Jalava et al. (2015)</td>
<td>Rank-based Reward</td>
<td>Symbolic Reward</td>
</tr>
<tr>
<td>Schreck (2020)</td>
<td>Relative Performance</td>
<td></td>
</tr>
</tbody>
</table>

systems. Because the organizational behavior studies are mostly field and lab experiments, they are therefore generally descriptive. In this paper, we build upon those descriptive studies and develop a prescriptive model to complement their findings. We borrow our terminology of “Relative” and “Absolute” rating systems from the organizational behavior literature (e.g., Goffin et al., 1996; Cascio and Aguinis, 2018). We next review some representative papers in the field of research on organizational behavior that have adopted similar rating systems to our conceptualization of Relative and Absolute systems.

Focusing on ratees’ behavior, Blume et al. (2009) show that ratees tend to be more attracted to a forced distribution system that has less stringent conditions for low performing groups (i.e., when the firm offers training opportunities for low performers rather than terminating their position with severance packages). Blume et al. (2013) further extends that study and show that ratees are more likely to be attracted to a forced distribution rating when they possess higher levels of cognitive ability. In addition, Berger et al. (2013) study ratee’s sabotage behavior and show that it has detrimental effect on implementing forced distribution rating. Considering both Relative and Absolute systems, Roch et al. (2007) show that workers perceive an Absolute rating system to be more fair and procedurally just than a Relative system. Similarly, in an educational context, Jalava et al. (2015) study how students performance vary between a rank-based system (when only the
top three students get an A) and a symbolic reward system (when students with a score above 90% receive a certificate), and show that the effects can vary by gender. That is, while girls get motivated by both systems, boys get motivated only by the rank-based system.

Focusing on raters’ behavior, Goffin et al. (1996) show that Relative rating has higher criterion-related validity than does the Absolute format. Similarly, Wagner and Goffin (1997) find that raters provide more accurate ratings under the comparative rating systems than Absolute rating systems in terms of differential elevation and stereotype accuracy. Considering a goal setting system, Fan and Gómez-Miñambres (2020) show that raters tend to set goals that are too challenging, resulting in sub-optimal performance.

Unlike the above studies that have focused on psychological traits of raters and ratees, our focus in this paper is on the rational decision processes of firms and workers. That is, we compliment this stream of research by comparing the performance of rating systems in equilibrium, and offer prescriptions as to how firms should implement rating systems based on the characteristics of jobs and rating pools to improve the overall performance of their workforce.

3. Model

We consider a rating pool that consists of $n \geq 2$ workers. The firm evaluates workers’ performance based on a Relative or an Absolute rating system. In both systems, each worker can receive either a high or a low rating. Workers who receive high ratings are deemed more ready for a promotion than others. Thus, we consider that receiving a high rating results in a value $B > 0$ for a worker (which can be non-monetary), while a low rating has no value to the worker.

The performance of worker $i$ is driven by his ability, $a_i$, and effort exerted, $e_i$, during the rating period. Ability is the collection of all knowledge and skills a worker brings to a job. Each worker knows his individual ability and the overall distribution of other workers’ abilities (i.e., a cumulative distribution function $F(a)$ and probability density function $f(a)$ with $a \in [a, \bar{a}]$). The effort, $e_i$, captures additional action, beyond the given ability, a worker exerts to improve his performance. This effort leads to an improvement $r(e_i)$ of the worker’s performance. This effort is not without cost as the worker incurs $c(e_i)$, which is increasing in $e_i$. The performance of a worker with ability
$a_i$ and effort level $e_i$ can be obtained as follows:

$$v_i(e_i) = a_i + r(e_i).$$  \hspace{1cm} (1)

This performance function preserves the heterogeneity of workers in a rating pool and the relationship between performance and efforts. This formulation also considers the substitutability of effort and ability, and accounts for the fact that workers can compensate for their lower abilities by exerting higher efforts. The workers cannot observe each other’s ability or effort. Similarly, the firm can only observe workers’ performance ($v_i(\cdot)$), but not their abilities and efforts. The workers receive high or low ratings based on their performance level and according to the adopted rating system (i.e., Relative or Absolute).

**Model of a Relative Rating System ($R$):** In a Relative system, workers receive high ratings according to the ordinal rank of their performance. We denote the allowable percentage of high ratings by $\rho \in (0, 1)$. This implies that the firm gives high ratings to the top $m = \lceil \rho n \rceil$ workers in the rating pool.

A worker with ability level $a_i$ chooses his equilibrium effort to maximize his individual payoff. Because worker $i$ does not know other workers’ ability levels, the equilibrium performance of other workers ($v^*$) is uncertain. We represent the probability of worker $i$’s performance being the $k$th highest of the $n$ workers by $P^m_k(e_i, a_i, v^*)$. Accordingly, worker $i$ chooses his equilibrium effort as follows:

$$e^R_i = \arg \max_{e_i} \sum_{k=1}^{m} B \cdot P^m_k(e_i, a_i, v^*) - c(e_i).$$  \hspace{1cm} (2)

In a symmetric Bayesian Nash equilibrium, all workers play the strategy $v^*(\cdot)$. Given the distribution of workers’ abilities, worker $i$ would outperform another worker $j$ if $v_i > v^*(a_j)$ (i.e., $a_j < v^{* - 1}(v_i)$), which is equivalent to $F((v^* - 1)(v_i))$ (recall that $F(a)$ is the cumulative distribution function of abilities). Accordingly, we obtain the probability $P^m_k(e_i, a_i, v^*)$ as follows:

$$P^m_k(e_i, a_i, v^*) = \frac{(n - 1)}{(k - 1)! (n - k)!} F(v^* - 1(v_i))^{n-k} (1 - F(v^* - 1(v_i)))^{k-1}. \hspace{1cm} (3)$$

The expected payoff of worker $i$ under the Relative system is $\pi^R_i = \sum_{k=1}^{m} B \cdot P^m_k(e^R_i, a_i, v^*) - c(e^R_i)$. 
In addition, the total expected performance of all workers under the Relative system is $V^R = K \int_{a_1}^{a_n} v_i^R \cdot f(a) \cdot da$, where $v_i^R = v_i(e_i^R)$ and $K$ captures the benefit to the firm per unit improvement in the expected performance of the rating pool.

**Model of an Absolute Rating System (A):** In an Absolute system, workers receive high ratings when their performance exceeds a standard threshold, denoted by $D$. This system could be viewed as being more predictable since each worker’s rating does not depend on the performance of other workers. A worker with ability $a_i$ chooses his effort to maximize his individual payoff as follows:

$$e_i^A = \arg \max_{e_i} B \cdot I\{v_i(e_i) \geq D\} - c(e_i).$$ (4)

The indicator function $I\{v_i(e_i) \geq D\}$ represents whether the worker’s performance is above or below the standard threshold. The expected payoff of worker $i$ under the Absolute system is $\pi_i^A = B \cdot I\{v_i^A \geq D\} - c(e_i^A)$ where $v_i^A = v_i(e_i^A)$. In addition, the total expected performance of all workers under the Absolute system is $V^A = K \int_{a_1}^{a_n} v_i^A \cdot f(a) \cdot da$.

**Firm’s and Workers’ Perspectives:** The firm’s objective is to maximize the total expected performance of workers in a rating pool. Specifically, the firm prefers an Absolute system if $V^A \geq V^R$, and it prefers a Relative system otherwise. As presented above, $V^A$ is determined by the workers’ equilibrium efforts under an Absolute system (i.e., $e^A = (e_1^A, ..., e_n^A)$), while $V^R$ is determined by the workers’ equilibrium efforts under a Relative system (i.e., $e^R = (e_1^R, ..., e_n^R)$). However, from the workers’ perspectives, they prefer a system that maximizes their individual payoffs. Specifically, worker $i$ prefers an Absolute system if $\pi_i^A \geq \pi_i^R$, and he prefers a Relative system if $\pi_i^R > \pi_i^A$.

In the next section, we characterize the workers’ equilibrium choices and compare the performance of the two systems. For analytical tractability, we consider $r(e_i) = \theta \ln(e_i+1)$ and $c(e_i) = ce_i$. A concave transformation of effort paired with a linear cost function is common in the literature (e.g., Terwiesch and Xu 2008; Körpeoğlu and Cho 2017; Rahmani et al. 2018). The parameter $\theta$ represents a worker’s efficiency of transforming effort to performance. The plus one inside the logarithmic transformation of effort ensures that when a worker exerts zero effort, he will perform at his given ability level (i.e., $v_i(0) = a_i$). Table 2 summarizes the key notation of this paper.
4. Results

4.1 Relative Rating System

In a Relative rating system, workers compete against each other for a limited number of high ratings. This implies that workers need to anticipate the equilibrium choices of other workers when choosing the amount of effort they are willing to exert to achieve a high rating. The next proposition characterizes the workers’ choices of equilibrium efforts under a Relative rating system.

**Proposition 1. Relative System Efforts:** Under a Relative rating system with $m$ high ratings, a worker with ability $a_i$ exerts the equilibrium effort $e_i^R$ such that

$$e_i^R = \frac{1}{c} \int_{\bar{a}}^{a_i} \exp \left( \frac{x - a_i}{\theta} \right) \cdot \sum_{k=1}^{m} B \cdot \frac{(n-1)!}{(k-1)!(n-k)!} \cdot f(x) \cdot H(x,n,k) \, dx \quad \forall i,$$

(5)

where $H(x,n,k) = F(x)^{n-k-1} (1 - F(x))^{k-2} (n-k-(n-1)F(x))$.

Proposition 1 shows that any worker with ability $a_i > \bar{a}$ has an incentive to exert some effort under a Relative system. Notably, even a worker with the highest ability in the pool chooses to exert effort (i.e., $e_i^R (a_i = \bar{a}) > 0$). The reason is that, in this system, having the highest ability does not ensure that the worker will receive a high rating, as other workers (with lower abilities)
can boost their performance by exerting high efforts.

Figure 1: Worker’s Effort under a Relative Rating System

\[ e_i^R(a_i) \]

Worker’s Ability \( (a_i) \)

\[ 0 \leq e_i^R(a_i) \leq 0.7 \]

Parameters: \( a_i \sim U[0,8] \), \( B = 1 \), \( \rho = 0.3 \), \( n = 10 \), \( \theta = 1 \), and \( c = 0.5 \).

Figure 1 illustrates how equilibrium efforts of workers depend on their abilities. Workers with low ability will exert the least amount of effort because of the low probability that they will achieve a performance high enough to be among the top \( m \) performers. Workers with mid-level abilities exert the highest efforts because of the competitive nature of the rating pool. In contrast, workers with high level abilities, who feel more confident about receiving a high rating, exert lower efforts to save costs.

Figure 2: Sensitivity of Workers’ Efforts under a Relative Rating System

(a) The Effect of Percentage Threshold \( (\rho) \)

\[ e_i^R(a_h) \]

Percentage Threshold \( (\rho) \)

(b) The Effect of Pool Size \( (n) \)

\[ e_i^R(a_l) \]

Pool Size \( (n) \)

Parameters: The same as in Figure 1 with \( a_l = 4 \) and \( a_h = 7 \).

Figure 2 shows the effect of the percentage threshold \( (\rho) \) and pool size \( (n) \) on the workers’ equilibrium efforts. Since the number of high ratings \( (m) \) depends on both parameters, each has a moderating effect on the workers’ efforts. Both figures display sharp changes in effort levels at values
where the combination of $\rho$ and $n$ results in an integer increase in the number of high ratings, $m$. Figure 2a shows that, as $\rho$ increases, workers’ efforts increase in a step-wise manner (based on integer increases in $m$) until the workers’ confidence in their chances of receiving high ratings becomes large enough that they reduce their effort levels (in a similar step-wise manner). The decrease in efforts happens sooner (at lower values of $\rho$) for workers with higher abilities than for those with lower abilities. The reason is that with a larger number of high ratings, higher ability workers can receive a high rating even with lower effort levels, whereas the low ability workers can get a high rating only if they exert a considerable amount of effort.

Figure 2b shows that the size of the rating pool can have different effects on workers’ efforts depending on their ability levels. First note that an increase in the pool size can have two effects on workers incentives to exert effort. On one hand, a larger pool size may increase the number of high ratings ($m$); on the other hand, it decreases the chances of each worker being among the top $m$ performers. As a result, the effect of the pool size can be different from the effect of the percentage threshold (as in Figure 2a). An increase in the pool size motivates workers with high abilities to exert higher efforts, while it demotivates workers with low abilities to exert effort. The reason is that, as the pool size increases, the likelihood of each worker outperforming other workers decreases. The workers with high abilities choose to compensate their lower chances of getting high ratings by exerting higher efforts. In contrast, workers with low abilities find it too costly to exert high efforts. This implies that as the pool size increases, the workers’ efforts, and in turn their performance, become more disperse.

4.2 Absolute Rating System

In an Absolute rating system, workers do not compete against each other. Instead, they are given a high rating if their performance meets or exceeds a standard threshold. Accordingly, each worker can choose his effort level based on the standard threshold and his ability. The next proposition characterizes workers’ choices of equilibrium efforts under an Absolute rating system.

**Proposition 2. Absolute System Efforts:** Under an Absolute rating system with a standard threshold $D$, a worker with ability $a_i$ exerts equilibrium effort $e_{iA}$ such that
\[ e^A_i = \begin{cases} 
0 & \text{if } a_i > a_{max} \\
\exp \left( \frac{D-a_i}{\theta} \right) - 1 & \text{if } a_{min} \leq a_i \leq a_{max}, \\
0 & \text{if } a_i < a_{min} 
\end{cases} \]

where \( a_{min} = \max \{ D - \theta \cdot \ln \left( \frac{B}{c} + 1 \right), \bar{a} \} \) and \( a_{max} = \min \{ D, \bar{a} \} \).

Proposition 2 shows that it is possible that only a fraction of workers choose to exert effort under an Absolute system (i.e., when \( a_{min} > \bar{a} \) or \( a_{max} < \bar{a} \)). In the case where \( a_{max} < \bar{a} \), a worker with \( a_i \geq a_{max} \) exerts no effort since his ability exceeds the standard threshold and he is guaranteed to receive a high rating. When \( a_{min} \leq a_i < a_{max} \), workers exert positive effort to perform up to the standard threshold. In addition, their effort is decreasing in their ability level, which is driven by the fact that ability and efforts are substitutable. When \( a_i < a_{min} \), workers do not exert effort because it is not cost effective for them to do so (i.e., their expected payoff becomes negative). In addition, as shown in equation (6), efforts increase as the standard threshold \( D \) increases. However, as \( D \) gets larger, the thresholds \( a_{min} \) and \( a_{max} \) may also increase, indicating that only higher ability workers choose to exert effort. Figure 3 illustrates these results.

Figure 3: Worker’s Effort under an Absolute Rating System

![Graph showing worker's effort under an Absolute Rating System](image)

Parameters: The same as in Figure 1 with \( \theta = 1.75 \) and \( c = 0.25 \).

### 4.3 Comparison of Relative and Absolute Rating Systems

In this section, we compare the two rating systems in terms of workers’ performance and payoff. In order to characterize closed-form solutions, we focus on a simple case with two workers (i.e., \( n = 2 \)).
We later generalize our results for the case with multiple workers in §5. In a Relative system with two workers, the number of possible high ratings is \( m \in \{0, 1, 2\} \). Note that if \( m = 0 \) or \( m = 2 \), workers have no incentives to exert effort in a Relative system. Hence, the performance of the Absolute system (weakly) dominates the performance of the Relative system. The next proposition characterizes the comparison of the two systems for the case where \( m = 1 \).

**Proposition 3. Comparison of Workers' Performance and Payoffs:** Suppose \( n = 2 \), \( m = 1 \), and \( a_i \sim U[0, 2b] \). There exist thresholds \( \hat{a} \) and \( \tilde{a} \) such that

(i) The performance of a worker with ability \( a_i \) is higher in an Absolute system than in a Relative system (i.e., \( v_i^A > v_i^R \)) if and only if \( a_{\text{min}} \leq a_i < \hat{a} \).

(ii) The payoff of a worker with ability \( a_i \) is higher in an Absolute system than in a Relative system (i.e., \( \pi_i^A > \pi_i^R \)) if and only if \( a_i > \tilde{a} \).

(iii) \( \hat{a} \geq \tilde{a} \). Both thresholds \( \hat{a} \) and \( \tilde{a} \) are non-increasing in \( B \) and they are non-decreasing in \( c \) and \( D \).

The first part of Proposition 3 shows that workers with average abilities perform better under an Absolute system than under a Relative system. The reason is that an Absolute system prompts these workers to perform up to the standard threshold, while a Relative system may not provide them with such incentives due to its competitive nature. In contrast, workers with high or low abilities (i.e., \( a > \hat{a} \) or \( a_i < a_{\text{min}} \)) perform better under a Relative system. These workers exert lower efforts under an Absolute system than a Relative system because either they can perform up to the standard threshold with a small amount of effort (when \( a > \hat{a} \)) or they do not find it cost effective to exert any effort (when \( a_i < a_{\text{min}} \)).

The second part of Proposition 3 shows that workers with high abilities prefer an Absolute system over a Relative system. These workers can receive high ratings with low or no effort under an Absolute system, while they need to exert higher effort under a Relative system. In contrast, workers with low abilities prefer a Relative system, because even with low effort, they have a chance of getting a high rating in a Relative system; whereas, they have to exert higher effort in an Absolute system to meet the standard threshold.

The third part of Proposition 3 shows that there are overlaps in the ranges of abilities where both workers and the firm prefer one rating system over the other. Figure 4 illustrates regions
of alignment and misalignment between the workers’ and firm’s perspectives. While high ability workers (i.e., with \( a_i \geq \bar{a} \)) prefer an Absolute system, their performance is higher in a Relative system, indicating a misalignment between the two perspectives. The workers with average abilities (i.e., with \( \bar{a} < a_i \leq \hat{a} \)) also prefer an Absolute system and their performance is also higher in an Absolute system, indicating an alignment between the two perspectives. In the region where \( a_{min} < a_i \leq \bar{a} \), workers perform better under an Absolute system since they exert higher effort to reach the standard threshold, but they prefer a Relative system because they can exert less effort and still maintain a high probability of receiving a high rating. Finally, low ability workers (i.e., with \( a_i \leq a_{min} \)) prefer a Relative system and also perform better under that system, because these workers do not find it cost effective to exert any effort in an Absolute system, and thus, they have no chances of receiving a high rating in that system. Figure 4 also illustrates that the region where the two perspectives align under an Absolute (Relative) system expands as the standard threshold \( D \) decreases (increases). The reason is that, a lower (higher) standard threshold encourages (discourages) low ability workers to exert effort and receive a high rating in an Absolute system.

We next compare the total expected performance of the workers under the two rating systems. Replacing equilibrium efforts (characterized in Propositions 1 and 2) in the total expected perfor-
Authors: The Implications of Rating Systems on Workforce Performance

Performance function of the workers (presented in §3), we obtain:

$$V^R = K \int^{\bar{a}}_2 \theta \cdot \ln \left[ \frac{1}{c} \int^{a_i}_2 \exp \left( \frac{x}{\theta} \right) \sum_{k=1}^{m} B \frac{(n-1)!}{(k-1)!(n-k)!} \cdot f(x) \cdot H(x,n,k)dx + \exp \left( \frac{a_i}{\theta} \right) \right] f(a_i) \, da_i, \tag{7}$$

$$V^A = K \int^{a_{\min}}_a x f(x) \, dx + K \int^{\bar{a}}_{a_{\max}} x f(x) \, dx + K \cdot D \cdot (F(a_{\max}) - F(a_{\min})), \tag{8}$$

where $H(x,n,k)$, $a_{\min}$, and $a_{\max}$ are as characterized in Propositions 1 and 2.

**Proposition 4. Comparison of Total Expected Performance:** Consider the total expected performance functions in equations (7) and (8).

(i) There exist thresholds $\rho$ and $\bar{\rho}$, such that $V^A > V^R$ if and only if $\rho < \rho$ or $\rho > \bar{\rho}$.

(ii) Suppose $a_i \sim U[0,2b]$. There exist thresholds $D$ and $\bar{D}$, such that $V^A > V^R$ if and only if $D < D < \bar{D}$.

Figure 5: Comparison of Total Expected Performance of Rating Systems

Parameters: The same as in Figure 1 with $\theta = 10.$

Proposition 4 shows that the comparison of the total expected performance of the two rating systems depends on the percentage threshold ($\rho$) or standard threshold ($D$), and Figure 5 illustrates this result. The overall takeaway from the proposition is that a rating system with a moderate number of high ratings leads to higher total expected performance. When $D$ is small (or $\rho$ is large), a large number of workers can receive high ratings even if they exert low efforts. In contrast, when $D$ is large (or $\rho$ is small), a small number of workers can receive high ratings but only if they exert high efforts. In such cases, since the chances of getting high ratings are small, workers have low
incentives to exert high efforts, which results in lower total expected performance. In the next section, we study how firms can choose percentage and standard thresholds to balance the number of high ratings and improve workforce performance.

4.4 Choices of Percentage and Standard Thresholds

As we discussed in §4.3, the comparison between the two rating systems depends on the percentage threshold ($\rho$) and standard threshold ($D$) under the Relative and Absolute rating systems, respectively. These thresholds determine the number of high ratings that the firm offers to its workers under each rating system. Some firms may have limited availability of high ratings (e.g., due to limited promotion slots), while some other firms may want to balance performance improvement with the number of high ratings their offer (e.g., when high ratings are associated with monetary rewards). We next present these two approaches for choosing thresholds of the two rating systems.

**Comparative Thresholds:** We consider situations where thresholds $\rho$ and $D$ are set such that the expected number of workers that receive high ratings are the same under both systems. The next proposition formalizes this result.

**Proposition 5. Comparative Thresholds:** For any percentage threshold $\rho$ in a Relative rating system, there exists a unique standard threshold $D^*(\rho)$ that ensures the expected number of workers that receive a high rating in an Absolute system is the same as in a Relative system. Specifically,

$$D^*(\rho) = F^{-1}(1 - \rho) + \theta \left[ \log \left( \frac{B}{c} + 1 \right) \right].$$  \hspace{1cm} (9)

Proposition 5 shows that a firm can limit the expected number of high ratings in an Absolute system by appropriately setting the standard threshold. Specifically, the threshold $D^*(\rho)$, which we refer to as the *comparative* threshold, results in the same expected number of high ratings as with the percentage $\rho$ in a Relative system. Accordingly, the comparative threshold is decreasing in $\rho$. Adopting a comparative threshold is especially useful in situations where the firm needs to limit the number of high ratings available.
Optimal Thresholds: We study situations where thresholds $\rho$ and $D$ are set such that they maximize the firm’s total expected payoff by accounting for the expected cost associated with offering high rating rewards. Specifically,

$$\rho^{**} = \arg \max_{0 \leq \rho \leq 1} V^R(\rho) - \rho \cdot B,$$

$$D^{**} = \arg \max_{D \geq 0} V^A(D) - r(D) \cdot B,$$

where $r(D) \equiv (1 - F(a_{\min}(D)))$ is the fraction of workers that receive high ratings in an Absolute system with a standard threshold $D$, and $V^R(\cdot)$ and $V^A(\cdot)$ are as presented in equations (7) and (8).

**Proposition 6. Optimal Thresholds:** The optimal thresholds are as follows:

(i) Under a Relative rating system, there exists a unique percentage threshold $\rho^{**}$ that maximizes the firm’s payoff. In addition, $\rho^{**}$ is decreasing in $c$ and it is increasing in $K$.

(ii) Under an Absolute rating system with $a_i \sim U[0, 2b]$, there exists a unique standard threshold $D^{**}$ that maximizes the firm’s payoff. In addition, $D^{**}$ is non-decreasing in $B$ and it is non-increasing in $K$ and $c$.

The first part of Proposition 6 shows that it is optimal for a firm to set a smaller percentage threshold under a Relative system when workers’ cost of effort ($c$) is high or the firm’s benefit from performance improvement ($K$) is low. When effort cost is high, workers have lower incentives to exert effort (as shown in Proposition (1)). By setting a low percentage threshold, the firm can motivate high ability workers to exert higher efforts due to the tighter competitive nature of the rating system (as shown in Figure 2a), which can in turn result in a higher total expected performance. Similarly, when the firm’s benefit from performance improvement is low, it is optimal to offer high ratings to a smaller number of workers (by setting a lower $\rho$) to save in the firm’s expected cost of offering high rating rewards.

The second part of the proposition shows that it is optimal for a firm to set a higher standard threshold under an Absolute system when workers’ reward ($B$) is high, their cost of effort ($c$) is low, or the firm’s benefit from performance improvement ($K$) is low. When workers’ reward is high, a larger fraction of workers choose to exert effort up to the standard threshold (i.e., $a_{\min}$ is smaller
as shown in Proposition (2)). Hence, the firm should set a higher standard threshold to not only improve the total expected performance, but also to limit the number of high rating rewards it offers. Similarly, when the firm’s benefit from performance improvement is low, it is optimal to set a higher $D$ (i.e., resulting in lower $r(D)$) to limit the number of workers who receive high rating rewards.

Overall, the analysis in this section shows that firms can use the number of high ratings offered as an additional lever to improve workforce performance. In §5.3, we conduct large-scale numerical analyses and compare the optimal number of high ratings under the two rating systems. Moreover, we examine how the optimal choices of thresholds (as opposed to comparative thresholds) affect the expected performance of the rating systems.

### 4.5 Choices of Rewards

In this section, we study situations where workers’ rewards are strictly monetary, and the firm can set them to maximize its total expected payoff. The next proposition formalizes this result.

**Proposition 7. Optimal Rewards:** The optimal rewards are as follows:

(i) Under a Relative rating system, there exists a unique reward $B^R$ that maximizes the firm’s total expected payoff. In addition, $B^R$ is decreasing in $c$ and it is increasing in $K$ and $\theta$.

(ii) Under an Absolute rating system with $a_i \sim U[0, 2b]$, there exists a unique $B^A$ that maximizes the firm’s total expected payoff. In addition, $B^A$ is non-decreasing in $K$.

Proposition 7 shows that there exist unique rewards that maximize the firm’s total expected payoff. Note that if there was no costs associated with offering rewards (e.g., when they are non-monetary), higher rewards are always preferred because workers’ efforts are increasing in rewards (Propositions 1 and 2). However, when rewards are monetary, the firm can choose the level that balances the cost of offering such rewards and improving the workers’ performances. In §5.4, we conduct large-scale numerical analyses and compare optimal rewards under the two rating systems and examine the effect of those on the overall performance of the rating pool.
5. Numerical Analyses

In this section, we focus on the rating system of the U.S. Army as a specific context to illustrate how our results can apply in practice. We first estimate values for model parameters using available data in §5.1, and then present our findings in §5.2-5.4.

5.1 Estimation of Parameters based on the Rating System of the U.S. Army

We estimate model parameters using the case of the U.S. Army where officers go through an annual evaluation process. In the Army’s system, commissioned officers are evaluated annually within rank-specific rating pools, where only a fraction of them can receive high ratings, referred to as Most Qualified reports. These ratings are key determinants of an officer’s future promotion to the next rank. We concentrate on officers in the rank of major working to obtain a high rating for promotion to the rank of lieutenant colonel to focus our analyses. We estimate the rating pool size based on the data from the U.S. Army Human Resources Command. As reported by Evans (2018), the median of the rating pool sizes in the rank of major is 15. Accordingly, we center our estimation at the median and consider $n \in [2, 28]$.

In order to estimate a value of achieving a high rating for a worker, we use the data on the salaries for active duty officers. The Defense Finance and Accounting Service (DFAS) publishes the annual salaries based on ranks (pay grades) and number of years of service (DFAS, 2021). The U.S. Army regulations require officers to have a minimum of 7.5 years of service before they can be considered for promotion to major and a minimum of 10.5 years of service before they can be considered for promotion to lieutenant colonel (Department of the Army, 2020). Using the pay table for 2021, we calculate the difference between annual salary ranges of the two ranks and find that the monetary values for promotion from major to lieutenant colonel can range from $18k to $35k (see Table A-1 in the appendix). To account for non-monetary values of receiving a high rating (more responsibility, higher stature, etc.), we estimate $B \in [18k, 70k]$.

We estimate the ability distribution of a rating pool of majors based on their years of service in that rank. Majors often change jobs and locations throughout their career. Consequently, majors with more years of service have an advantage over less experienced ones because they are more familiar with the locations, people and systems necessary to efficiently execute their jobs. According
to Title 10 of the United States Code, majors can remain in their rank for up to ten years (United States Code, Title 10, 2018). Thus, we estimate that at the time of an evaluation, a major may have less than one year to at most 10 years of experience in that role. In the Army’s system, some rating pools may have small spreads of ability (e.g., in units that are newly formed or have a high turnover rates) and some may have large spreads of abilities (e.g., in special operations units and the West Point faculty). To capture the differences in the spread of abilities, we estimate the mean of the ability distribution at $b \in [0.5, 5]$ with $a_i \sim U [0, 2b]$. Although we normalize the minimum ability to zero, we note that similar results hold when we shift the distribution to the right.\(^2\)

We estimate the benefit per unit improvement in the expected performance of the rating pool by using the data on the U.S. Army’s revenue and personnel cost. Using the data on the annual congressionally approved funds that the U.S. Army receives from the government for procurement, research, and development, and the data on the U.S. Army’s working capital funds from the sale of goods and the provision of services for a fee to its customers, we estimate the revenue ranges between $49.8$ and $55.8$ Billion from FY 2018 through 2020. Similarly, using the data on the annual congressionally approved funds that the U.S. Army receives for regular Army personnel (U.S. Army Financial Management and Comptroller, 2021), we obtain that the personnel cost ranged between $40.9$ and $42.7$ Billion from FY 2018 through 2020. (See details in Table A-2 in the appendix.) Taking the ratio of the Army’s revenue to the personnel costs, we obtain that the Army’s revenue is on average 25% higher than its cost of personnel. In order to account for other intangible benefits of workforce, we consider the rate to be 50% in our estimation. That is, we estimate that the net value of each regular Army workforce is on average 1.5 times of their salary. Using the pay table for 2021, we calculate the range of salary increase for majors per additional year of experience (see Table A-1 in the appendix.). Then, as a proxy for the benefit per unit improvement in an officer’s performance, we multiply the range of salary increase by 1.5, and estimate $K \in [1.4k, 4.4k]$.

We next estimate the unit cost of effort by using the data on majors’ pay rate and working hours. Since majors in the same rating pool are typically subject to the same environmental conditions (location, responsibility, etc.), the differences in their pay rates can be a good proxy for their opportunity cost of improving performance (e.g., a major with 8 years of service can exert

\(^2\)For instance, suppose we consider $a_i \in [l, 2b + l]$. Using change of variables, we can rewrite the performance function in equation (1) as $v_i = l + \hat{a}_i + rz(e_i)$ where $\hat{a}_i \in [0, 2b]$. Thus, when comparing the performance of two systems, the constant term $l$ cancels out, and we obtain similar results.
additional effort to perform as high as a major with 10 years of service, while is being paid based on 8 years of service). Using the data from the Office of Personnel Management (OPM, 2021) and the Defense Finance and Accounting Service (DFAS, 2021), we find the annual pay rate for a major ranges from $86.3k (for over 8 years of service) to $101.8k (for 16 years of service and beyond). Thus, the average opportunity cost of improving performance by one unit is about $1.93k (i.e., $(101.8 - 86.3)/8$). Depending on tasks and responsibilities, officers may work a minimum of 5 days per week and 8 hours per day and a maximum of 7 days per week and 24 hours per day (e.g., when they are expected to be available for work 24/7 while deployed). This suggests that the difficulty of the job can be up to 4.2 times higher than the minimum (i.e., $(24 \times 7)/(5 \times 8) = 4.2$). Hence, considering $1.93k$ as the base cost of effort, we estimate $c \in [1.9k, 8.1k]$. We normalize the cost to efficiency ratio ($\delta \equiv c/\theta$) between zero and one, and estimate $\delta \in (0, 1]$. Table 3 summarizes the ranges of parameters for our numerical study.

Table 3: Parameter Ranges (all drawn from Uniform distributions)

<table>
<thead>
<tr>
<th>Parameter</th>
<th>$n$</th>
<th>$K$</th>
<th>$B$</th>
<th>$b$</th>
<th>$c$</th>
<th>$\delta$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Range</td>
<td>$[2, 28]$</td>
<td>$[1.4k, 4.4k]$</td>
<td>$[18k, 70k]$</td>
<td>$[0.5, 5]$</td>
<td>$[1.9k, 8.1k]$</td>
<td>$(0, 1]$</td>
</tr>
</tbody>
</table>

To gain a comprehensive understanding of the comparison between the two rating systems, we conduct extensive numerical analyses where we compare the total expected performance of the rating pool under the two systems while considering comparative thresholds, optimal thresholds, and optimal rewards. The equilibrium efforts characterized in Propositions 1 and 2 allow us to solve the problems efficiently. In each run of the numerical experiment, problem parameters are drawn randomly from their respective uniform distribution, presented in Table 3. This sampling approach allows us to derive insights that would be structurally similar to an exhaustive multi-dimensional computation. This approach is consistent with recommendations in Nelson (2013) and has been used in prior studies (e.g., Roels and Tang 2017; Zorc et al. 2017; Rahmani and Ramachandran 2020).

5.2 Comparison of the Rating Systems with Comparative Thresholds

We randomly generated 1000 sets of parameters within the ranges shown in Table 3. The U.S. Army currently employs a Relative rating system that requires the percentage threshold ($\rho$) to be
less than 50 percent (Department of the Army, 2019). To examine the impact of more or less restrictive percentages on the comparison between the two rating systems, we center our estimation of \( \rho \) at 0.5 and consider \( \rho \in [0.2, 0.8] \). Naturally, we only consider parameter sets that would comprise feasible instances where the firm would award a high rating to at least one worker and not all workers (i.e., \( 1/n \leq \rho < 1 \)). We discarded infeasible instances (which were less than 3% of the samples) and redrawn them to achieve 1000 feasible sets of parameters. In addition, we calculated the comparative threshold of the Absolute system \( (D^*(\rho)) \) as presented in Proposition 5, which ensures that both systems result in the same expected number of high ratings.

Table 4 shows the percentage of cases that the identified system outperformed the other with comparative thresholds. The Relative system resulted in a higher total expected performance in 92.5% of the instances and the Absolute system resulted in a higher total expected performance in the remaining 7.5% of the instances. In addition, the performance gain of the Relative system when it is optimal is on average 8.57%, while the same for the Absolute system is on average 8.37%.

Table 4: Comparison of the Total Expected Performance of Relative and Absolute Systems (i.e., \( V^R \) and \( V^A \)) with Comparative Thresholds

<table>
<thead>
<tr>
<th>System</th>
<th>% Optimal</th>
<th>% Gain</th>
</tr>
</thead>
<tbody>
<tr>
<td>Relative</td>
<td>92.50%</td>
<td>8.57%</td>
</tr>
<tr>
<td>Absolute</td>
<td>7.50%</td>
<td>8.37%</td>
</tr>
</tbody>
</table>

We also compare the total expected payoff of the rating pool under the two rating systems. Specifically, we examine \( \Pi^S = \int_2^\pi (a + r(e^S(a)) - c(e^S(a))) f(a) da \) for \( S \in \{ R, A \} \). We obtain that the total expected payoff is higher in the Absolute than in the Relative system (i.e., \( \Pi^A \geq \Pi^R \)) in all of the instances. The reason is that, while both systems result in the same expected number of high ratings (given the comparative thresholds), the total cost of efforts is lower in the Absolute system, as only a fraction of workers may exert effort (see Propositions 1 and 2).

To decipher the circumstances under which a firm may prefer one system to the other, we present how various model parameters affect the total expected performance of the two systems in Table 5. This table was created by averaging the values for each particular parameter when Relative or Absolute systems were optimal. Table 5 shows that an Absolute system performs better when the workers’ cost of effort relative to their efficiency rate is low (i.e., \( \delta \) is low), as in many routine jobs. In contrast, a Relative system performs better when the cost of effort relative to efficiency rate is
high, as in many knowledge-intensive jobs. In the case of the Army, the cost of effort relative to efficiency could be high in operationally deploying units, because of the amount of time soldiers spend training in unfamiliar places or the uniqueness of the tasks they are asked to accomplish. Conversely, training and garrison units that operate the Army’s posts have schedules that are very cyclical in nature and the repetitiveness of their tasks likely leads to lower cost of effort relative efficiency rate.

Table 5: The Impact of Model Parameters on the Performance of Rating Systems

<table>
<thead>
<tr>
<th>System</th>
<th>% Optimal</th>
<th>$b$</th>
<th>$B$</th>
<th>$n$</th>
<th>$\rho$</th>
<th>$c$</th>
<th>$\theta$</th>
<th>$\delta$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Relative</td>
<td>92.50%</td>
<td>2.79</td>
<td>45.32</td>
<td>16.12</td>
<td>0.51</td>
<td>4.91</td>
<td>12.17</td>
<td>0.56</td>
</tr>
<tr>
<td>Absolute</td>
<td>7.50%</td>
<td>2.41</td>
<td>36.11</td>
<td>6.40</td>
<td>0.52</td>
<td>6.07</td>
<td>18.89</td>
<td>0.45</td>
</tr>
</tbody>
</table>

Moreover, Table 5 shows that when an Absolute system is optimal, the average size of the rating pool is smaller than when a Relative system is optimal (i.e., 6 as opposed to 16). Indeed, we find that when the pool size is less than 10 (which comprises nearly 30% of our samples), the percentage of cases where an Absolute system is optimal triples to 22.7%. While workers’ efforts under an Absolute system do not depend on the size of the rating pool (Proposition 2), the pool size significantly impacts the efforts and performance of workers in a Relative system. As shown in Figure 2b, the change in the pool size can have opposing effects on the workers’ choices of efforts under a Relative system. However, as $n$ gets smaller, the negative effect on efforts of high ability workers dominates the positive effect on efforts of low ability workers. In the case of the Army, more than 40% of the rating pool sizes for majors are less than 10 (Evans, 2018). These findings suggest that the Army can improve the performance of small rating pools, especially in training and garrison units (where $\delta$ is also low), by considering an Absolute rating system. Such an approach not only enhances the expected performance of those rating pools, but also improves the workers’ total expected payoffs.

5.3 Comparison of the Rating Systems with Optimal Thresholds

For all 1000 randomly-generated sets of parameters within the ranges shown in Table 3, we calculated optimal thresholds $D^*$ and $\rho^*$, as presented in Proposition 6. In all cases, the results yielded feasible instances where some portion of workers exert effort in each system (i.e., $1/n \leq \rho^* < 1$
and $a_{\min}(D^{**}) < a_{\max}$). We denote the optimal number of high ratings in a Relative system by $m^R$ (which is equivalent to $\lfloor n \cdot \rho^{**} \rfloor$) and in an Absolute system by $m^A$ (which is equivalent to $\lfloor n \cdot (1 - a_{\min}(D^{**})/2b) \rfloor$).

The analysis shows that the Relative system results in a higher optimal number of high ratings (i.e., $m^R > m^A$) in 21% of the instances and the Absolute system results in a higher number of high rating (i.e., $m^R \leq m^A$) in the remaining 79% of the instances. These results indicate that firms can use the number of high ratings as an additional lever to improve workers’ performance under each system.

<table>
<thead>
<tr>
<th>Performance/# of High Ratings</th>
<th>$m^R &gt; m^A$</th>
<th>$m^R \leq m^A$</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>$V^R &gt; V^A$</td>
<td>20.30%</td>
<td>67.10%</td>
<td>87.40%</td>
</tr>
<tr>
<td>$V^R \leq V^A$</td>
<td>0.70%</td>
<td>11.90%</td>
<td>12.60%</td>
</tr>
<tr>
<td>Total</td>
<td>21.00%</td>
<td>79.00%</td>
<td>100%</td>
</tr>
</tbody>
</table>

Table 6 shows the percentage of cases that the identified system outperformed the other with optimal thresholds. The Relative system resulted in a higher expected performance in about 87% of the instances and the Absolute system resulted in a higher expected performance in nearly 13% of the instances. In addition, we find that in the majority of the cases where a Relative system results in a higher total expected performance, the expected number of high ratings is lower under a Relative system than under an Absolute system. In contrast, in the majority of the cases where an Absolute system results in a higher total expected performance, the expected number of high ratings is higher in an Absolute system than in a Relative system. These results imply that the firm should set the number of high ratings in a way that exploits the difference between the two rating systems. That is, under an Absolute rating system, the firm should set a low standard threshold to encourage low ability workers to exert high efforts, whereas under a Relative system, it should set a small percentage threshold to encourage high ability workers to exert high efforts.

We also compare the total expected payoff of the rating pool under the two rating systems with optimal thresholds. We obtain that the total expected payoff is higher in the Absolute system than in the Relative system (i.e., $\Pi^A \geq \Pi^R$) in nearly 99.50% of the instances, which is due to the fact that the optimal number of high ratings is higher in the Absolute system in the majority of the cases (as shown in Table 6). These findings suggest that although firms can maximize the total expected
performance of the rating pool by optimally setting thresholds under a Relative system, that would not significantly improve the misalignment between the firm’s and workers’ perspectives.

5.4 Comparison of the Rating Systems with Optimal Rewards

For all 1000 randomly-generated sets of parameters within the ranges shown in Table 3, we calculated optimal rewards ($B^R$ and $B^A$), as presented in Proposition 7. In order to focus on the pure effect of rewards on performance, we considered comparative thresholds (as in Proposition 5). Similar to our previous analyses, we focus on feasible instances where some portion of workers exert effort in each system (i.e., $B^R > 0$ and $B^A > 0$). We discarded infeasible instances (which were less than 16% of the samples) and redrawn them to achieve 1000 feasible sets of parameters.

We find that in all cases a Relative system results in a higher expected performance than an Absolute system when considering optimal rewards (i.e., $V^R > V^A$). In addition, in 99.6% of the instances, the magnitude of the optimal reward under the Relative system exceeds that of the Absolute system (i.e., $B^R \geq B^A$). Specifically, we find that on average, the magnitude of the optimal reward under the Relative system is 49.1% higher than the same under the Absolute system. The implication of these findings is that although optimal rewards overwhelmingly favor the use of a Relative system, a firm with limited budget may not be able to extract the full benefits of implementing a Relative system.

We also compare the total expected payoff of the rating pool under the two rating systems. We obtain that the total expected payoff is higher in a Relative system than in an Absolute system (i.e., $\Pi^R \geq \Pi^A$) in nearly 44.6% of the instances. The reason is that the optimal reward under a Relative system is considerably higher than under an Absolute system. This implies that when the firm has an ample budget to optimally set rewards under the Relative system, it can maximize the total expected performance of the rating pool and also improve the alignment between its perspective and the workers’ perspective.

6. Conclusions

Firms institute rating systems to motivate workers to perform at their highest level. A Relative rating system compares the performance of workers against their peers and uses the effect of com-
petition in the workplace to incite workers to increase performance. In contrast, an Absolute rating system awards high ratings to workers when their performance meets or exceeds a standard threshold. In this paper, we study the implications of these rating systems on workers’ performance, and demonstrate several insights on when and why each of these rating systems can be beneficial.

We find a Relative rating system prompts all workers to exert effort, whereas an Absolute rating system may prompt only a fraction of workers to exert effort. However, a Relative system provides lower incentives for workers with intermediate abilities to exert effort as compared to an Absolute system. Accordingly, we find an Absolute system can result in a higher expected performance than a Relative system. This specifically happens when the rating pool is small and the job is routine (i.e., cost of effort relative to efficiency rate is low); otherwise, a Relative system results in a higher expected performance. In addition, a Relative system outperforms an Absolute system when the percentage of workers who can receive high ratings is neither too small nor too large. The reason is that a small (large) percentage threshold demotivates workers with low (high) abilities to exert high efforts, while a moderate percentage threshold motivates a larger fraction of workers to exert high efforts. When a firm can alter the rating thresholds to maximize performance, we find that the firm should set a low standard threshold under an Absolute rating system to encourage low ability workers to exert efforts, but it should set a small percentage threshold under a Relative system to encourage high ability workers to exert high efforts. Finally, when considering optimal rewards, we find that by offering higher rewards, the firm can improve the total expected performance of the workforce under both systems, but the effect is more pronounced under a Relative system than an Absolute system.

We also identify conditions where there is an alignment (or misalignment) between the firm’s and workers’ preferences on rating systems. We find that higher ability workers prefer an Absolute system due its predictable nature, while lower ability workers prefer a Relative system as it offers them a higher chance of receiving a high rating. Not only do low ability workers prefer a Relative system, their performance is also higher in a Relative system than in an Absolute system, indicating an alignment between the firm’s and workers’ perspectives. In contrast, high ability workers, who prefer an Absolute system, have higher performance in a Relative system, indicating a misalignment between the firm’s and workers’ perspectives. We show that firms can enhance the alignment between their perspective and workers’ perspective by optimally setting thresholds and/or rewards.
While setting optimal rewards significantly enhances the alignment between the two objectives, setting optimal thresholds only marginally reduces the gap between the two perspectives.

The United States Army serves as a good example for the analysis that we have put forth in this paper. Currently, senior officers and non-commissioned officers in all units of the Army are rated according to a Relative rating system where the top 49% and 24%, respectively, of a rank-specific rating pool can receive a high rating. In operationally deploying units, the cost of effort might be high relative to efficiency because of the amount of time soldiers spend training in unfamiliar places or the uniqueness of the tasks they are asked to accomplish. From our analysis, these types of units might benefit from using a Relative system to award high ratings, especially when the pool sizes are large. Conversely, training and garrison units that operate the Army’s posts have set schedules that are cyclical, and the repetitiveness of their tasks could yield a lower cost of effort relative to efficiency. Our study shows that these types of units may benefit from adopting an Absolute rating system to award high ratings for promotion. In addition, because of the rigidity of the Army’s personnel structure and compensation systems, such units can benefit from establishing comparative standard thresholds based on the possible number of high ratings.

This study offers several opportunities for future research to provide further insights in comparing the performance of Relative and Absolute rating systems. First, the most fruitful and promising research direction could involve an empirical or experimental assessment of the impact of rating systems on workforce performance. Such a study can also capture situations where a firm is unable to perfectly observe workers’ performance or when workers are risk-averse. For instance, future research can build on this study by considering an additional level of uncertainty (e.g., noise) in the performance function of the workers. Although we conjecture that such a noise factor may not have a significant effect on the comparison of the two systems, exploring the effect of that on workers’ choices of efforts under each system could be insightful. Similarly, future research can consider situations where workers are risk-averse. While intuition suggests that risk-averse workers would have stronger preference for the Absolute system, exploring the effect that on firm’s preferences can be informative. Second, the study can be extended to a multi-period rating system, where performance evaluation of workers occur multiple times during their employment at a firm, and their ratings can be accumulated for future promotions. We conjecture that workers would exert higher efforts in later rounds of evaluation under both systems, but the effect could be stronger in
a Relative system than in an Absolute system. Finally, it would be worthwhile to study situations where the firm is unable to evaluate all workers at the same time and offers high ratings sequentially. For instance, this arises when workers go through the evaluation process based on the date they were hired and not a fiscal year review cycle. We conjecture that such situations may not significantly affect workers’ equilibrium choices of efforts, but they can have implications for the firm’s optimal choices of thresholds and rewards. We hope future research can build on this study to refine our insights by exploring how the above scenarios affect the effectiveness and comparison of evaluation rating systems.

References


33