

Performance Feedback and Productivity: Evidence from a Field Experiment

Amrou Awaysheh
Kelley School of Business
Indiana University
801 W Michigan St
Indianapolis, IN 46202
awaysheh@iu.edu
+1 (317) 274 0955

Rocio Bonet
Human Resource Management Department
IE University
Maria de Molina 11, 13, 15
28006 Madrid, Spain
Rocio.bonet@ie.edu
+34 915 689 843 (Ext. 49843)

Jaime Ortega *
Department of Business Administration
Universidad Carlos III de Madrid
Calle Madrid 126
28903 Getafe (Madrid), Spain
jortega@emp.uc3m.es
+34 916 249 612

* corresponding author

History of Submission: First Submitted: May-2021, Major Revision: July-2021, Second Submission: January-2022, Minor Revision: March-2022, Third Revision: June-2022, Conditional Accept: June-2022, Fourth Submission: July-2022, Accept: July-2022.

Forthcoming at the *Production and Operations Management Journal*.

Performance Feedback and Productivity: Evidence from a Field Experiment

Abstract

We theorize that employees use the performance feedback they receive to reassess their beliefs about the marginal benefit of their effort, which may lead them to increase or reduce their effort. To test our model, we conduct a field experiment at the distribution center of a Fortune 500 firm where employees receive individual performance pay, and we study two types of feedback, individual and relative. The results show that employees react to feedback content in a way that is consistent with the model: they increase their effort if the information provided implies that the marginal benefit of increasing effort is high and decrease it if they learn that it is low. Moreover, performance feedback has a greater impact on the lower quantiles of the distribution of productivity.

Keywords: performance feedback, feedback content, productivity, incentive pay, information systems, distribution center

Received: May 2021; accepted: July 2022 by Anil Arya after two revisions.

1. Introduction

Organizations are increasing the amount and frequency of the performance feedback they provide to their employees (Ewenstein et al. 2016; Cappelli and Tavis, 2016; 2018). This surge is based on the generalized belief that supplying workers with ongoing feedback can boost their performance. Accordingly, companies are replacing their traditional performance appraisal practices, in which they give performance feedback once per year, with ongoing feedback systems as exemplified by the touchpoint or check-in systems used by GE, Deloitte, or Adobe¹ (Buckingham and Goodall 2015; Baldassarre and Finken 2015). These changes have been facilitated by developments in information technology systems that have reduced the costs of tracking individual performance, comparing the performance of different workers, and sharing this information with employees (Aral et al. 2012).

The information that employees receive about their own performance can motivate them to work better through different mechanisms. When individuals are paid according to their performance, feedback may tell them where they stand within the incentive scheme of the organization, and they may use this information to improve their individual performance. Thus, feedback can work as a complement to incentive pay (Lourenço 2016). In addition, receiving relative performance feedback can affect employees'

¹ <https://www2.deloitte.com/insights/us/en/focus/human-capital-trends/2015/performance-management-redesign-human-capital-trends-2015.html>, retrieved on March, 24, 2021.

self-esteem² (Kuhnen and Tymula 2012), which can also lead to a change in performance: as self-esteem is determined by relative standing among peers, employees who learn that they are doing poorly relative to others may want to increase their performance to improve their self-esteem. It is also possible that, when it is difficult for employees to tell whether a given level of performance is high or low, relative performance feedback may help them assess their own performance level (Kolstad 2013). This may also lead employees to change their behavior if they realize, for instance, that their performance was low.

However, although performance feedback could theoretically lead to an improvement in performance, empirical research has found mixed effects. This has led scholars across different disciplines to develop new research perspectives to understand the theoretical explanations behind the heterogeneous effects of performance feedback (Kluger and DeNisi 1996; Sprinkle 2000, 2003). For example, in the psychology literature, Kluger and DeNisi (1996) provide a comprehensive theory that explains variation in the performance consequences of feedback as a function of how a given feedback intervention shifts attention between the task and the self. Feedback interventions that shift the locus of attention towards the task should have stronger effects than those that shift attention away from it and towards the self. More recently, accounting scholars have theorized and found that the performance consequences of feedback vary with the characteristics of the feedback (e.g., Casas-Arce et al. 2017; Hannan et al. 2018), the context in which it is given (e.g., Tafkov 2013; Hannan et al. 2008) and the characteristics of the receiver (e.g., Lourenço et al. 2018). Most closely related to our paper, researchers have also explored the role played by the “content” of the information (e.g., Casas-Arce and Martinez-Jerez 2013, Kolstad 2013, Lourenço et al. 2018), i.e. the specific information received by each employee.³

In this study, we contribute to the research on the role of the content of feedback. We propose a formal moral hazard model in which employees choose the amount of effort based on their beliefs about their marginal product of effort, which is not completely known to them. Employees’ beliefs are based on their own information and on the information provided by the firm. Thus, when the firm provides feedback, employees’ beliefs may change, leading to a change in effort. The model provides two results. First, we show that feedback may increase or reduce worker effort. Second, we show that the size of the change in effort depends on the content of the feedback. Specifically, a content that indicates that the marginal benefit of effort is more likely to be high will lead to a greater increase in effort. Quite importantly, this prediction

² Self-esteem has been defined as an individual’s evaluation of him or herself (see e.g. Coopersmith 1967, pp. 4-5). According to the self-evaluation maintenance (SEM) model (Tesser 1988), individuals behave in order to maintain or improve their self-evaluations, which are influenced by the performance of “close others”. A more positive self-evaluation has a positive effect on affect, which is consistent with the utility function assumed here.

³ We use the term “content” to refer to the information received by each individual. Thus, consider a feedback policy in which employees are told their previous week performance. The “content” would be the specific value of performance that each employee is told. Although the feedback policy is the same for all employees, the content of the feedback varies across employees because their performance also varies.

does not rely on workers' pre-feedback beliefs and therefore its empirical test requires knowing which information was told to the workers but not what workers initially believed.

Our model complements other formal models of feedback which have also analyzed how workers' responses to feedback vary according to its content, based on the idea that workers learn from the feedback and change their behavior according to the new information. Such models have been proposed in contexts like tournaments (Casas-Arce and Martínez-Jerez 2009) and settings in which conformity to social norms is important (Chen et al. 2010). Our model instead analyzes responses to feedback when employees receive individual performance pay and social norms are not important. There are also models that have analyzed how much information is worth revealing, i.e., what the optimal amount of feedback is (Ertac 2005; Azmat and Iriberry 2010) and the effects of the quality of the information (Kolstad 2013) and the detail and frequency of the feedback (Casas-Arce et al. 2017) on workers' reactions. However, these models do not address the question of why responses to feedback are heterogeneous across employees who are subject to the same feedback policy. Our model instead provides predictions on how effort will vary according to the content of the feedback, thus providing an explanation for the heterogeneous effects of feedback. Our model is most closely related to Kuhnen and Tymula (2012), whose model also considers how effort changes as workers' beliefs change. The main difference with respect to their model is that in Kuhnen and Tymula (2012) the learning process itself is not modelled, while in our case learning is endogenous and constitutes the central part of the model.

To test the predictions of the model we conducted a field experiment⁴ at a warehouse distribution center of a Fortune 500 company. In the experiment we manipulated the type of feedback given to employees. The feedback could be individual (i.e., about the worker's own performance) or relative (i.e., including information on performance relative to coworkers). We used a 2x2 design and randomly assigned each of the four sections of the facility to one of the following treatments: individual feedback, relative feedback, both, and a control group. Because our purpose was to investigate employees' reactions to different pieces of information, each type of feedback treatment was designed in such a way that employees would receive information that was relevant to learning about the benefits of increasing performance. The feedback that each employee received referred to their performance in the previous week, and the feedback was given privately.

The rest of the paper is organized as follows. In section 2 we review the current literature on the heterogeneous effects of feedback and in section 3 we develop our formal model. In the following two sections we describe the warehouse in which we conducted the experiment (section 4) and the experiment

⁴The field experiment was approved by the institutional review board of the institution of the coauthor who conducted it.

itself (section 5). In section 6 we describe the data and variables, in section 7 we describe the econometric results and in section 8 we discuss our findings.

2. Related literature

In recent years scholars have made some progress to understand why the effects of feedback on performance are heterogeneous, focusing on aspects such as the characteristics of the feedback, the context in which feedback is given, the receiver of the feedback, and the information content of the feedback. In this section, we summarize the main empirical findings of this literature (see Table 1 in e-companion for a summary of the relevant empirical literature).

Regarding the characteristics of the feedback, researchers have considered how the effects of feedback vary depending on its level of detail (Hannan et al. 2018; Casas-Arce et al. 2017), the frequency with which it is given (Casas-Arce et al. 2017) and whether it is provided in a private or public way (Hannan et al. 2013; Tafkov 2013; Casas-Arce et al. forthcoming). Scholars have also explored how the effects of feedback vary with the context in which it is given. For example, the type of pay scheme (Tafkov 2013; Hannan et al. 2008) or a tournament's prize structure (Newman and Tafkov 2014) have been found to be important factors in determining the effectiveness of feedback. Another contextual factor that matters is the ability of workers to choose where to allocate effort in a multitask environment (Hannan et al. 2013). Some studies have also found that individuals react differently to feedback because they have unique characteristics, such as distinct attitudes toward feedback (Lourenço et al. 2018), varying levels of experience in the organization (Blanes i Vidal & Nossol 2011), or different levels of ability and education (Casas-Arce et al. forthcoming).

Of particular relevance for our study is the growing research that has tried to explain the heterogeneous effects of feedback as a function of its content, i.e., the information received by each individual with the feedback intervention. How much employees learn from the feedback may vary across individuals if their pre-feedback beliefs about their performance are to some extent biased and these biases vary across individuals: those with more biased beliefs should react more to feedback. In particular, individuals may react differently depending on whether they learn that they were overestimating or underestimating their performance (Lourenço et al. 2018; Azmat et al. 2019; Kuhnen and Tymula 2012; Chen et al. 2010; Jung et al. 2021).

Learning is also heterogeneous across individuals because feedback provides personalized information, which varies across employees. When they receive feedback employees may update their beliefs about their marginal benefit of exerting more effort, and those who learn that the marginal benefit of effort is smaller than they believed may react by reducing performance, while those who learn that the marginal benefit is higher may react by increasing performance. An important determinant of the marginal benefit of effort is

where the individual stands in terms of performance prior to receiving feedback (Casas-Arce and Martinez-Jerez 2013; Eyring and Narayanan 2018). This has been explored in the context of relative feedback and of individual performance feedback. For example, Casas-Arce and Martinez-Jerez (2009) found that providing sales agents with interim feedback about how they stand with respect to others led to decreased effort for those who were ahead of their peers and for those trailing behind who determined that they had a very large performance gap. In an experiment with a sample of students, Eyring and Narayanan (2018) also found that higher reference points that are farther away from an individual's performance (i.e., top quartile for low performers) may lead to lower performance because they offer a lower marginal utility of effort. In other words, when learning about their performance relative to the top quartile, poor performers may have low expectations of reaching such high standards, which may discourage them from trying harder. In contrast, using a quantile regression approach, Azmat and Iriberry (2010), in a high-school context, found that informing students about their relative performance (above or below average) as well as the distance from this average resulted in an increase in performance throughout the whole performance distribution. Lourenço et al. (2018) explored the performance effects of giving individual feedback to physicians in the context of negative incentives (i.e., termination threshold) and found that feedback improved effort both less and later for poor performers (i.e., those under termination threat). However, the effects were not observed for high performers. In a related study, not considering performance but the revision of their goals, Ilies and Judge (2005) found that those who were performing above their goals reacted to positive feedback by engaging in upward goal revision and those who were performing below their goals reacted to negative feedback by lowering their goals—a finding suggesting that individuals who see themselves far from their goals may stop trying. The conclusion of these studies is that personalized messages that inform individuals about the marginal benefits of increasing effort may lead to heterogeneous behavioral responses. Our study extends this literature and proposes a model to explain the heterogeneous impact of feedback depending on the information provided about the marginal benefits of effort.

While the idea that the content of feedback affects performance has appealed to several researchers, only a few studies have provided empirical estimates of the content per se, i.e., have included in their estimations the actual content of the information given (i.e., the *values* of the variables that constitute the feedback given to each employee). For example, in a laboratory experiment, Gill et al. (2019) provided an estimate of how the actual content of the rank (i.e., the rank-order feedback) impacted the effort provision of the individuals who received the information. Casas-Arce and Martinez-Jerez (2009) provided estimates of the effects of the actual leading and trailing distances between the rankings of the players in the contest. Kolstad (2013) estimated the effect of the amount of new information given by the feedback as the distance between the adjusted risk performance and the expected performance of the surgeon prior to the feedback. Other studies classified the content of the feedback as either “positive” or “negative” based on whether the

feedback implied that employee performance was high or low, respectively, and estimated two treatment effects of feedback, one for employees who received positive feedback and another for those who received negative feedback (e.g., Casas-Arce et al. 2017; Chen et al. 2010).

Our paper contributes to this literature on feedback content. In our study, we hypothesize that a possible reason why feedback content matters is that employees may use the information that they receive to update their beliefs about the marginal benefit of exerting effort. Thus, we propose a specific mechanism to explain why employees who receive different information may react differently to the feedback. From a theoretical standpoint, our main contribution is to provide a formal model of how changes in content influence changes in beliefs, which in turn influence worker effort. Our model extends the work of Kuhnen and Tymula (2012), whose model considers changes in workers' beliefs as exogenous. Our model instead considers how changes in beliefs are driven by changes in feedback content. Another important contribution of our model is to provide a testable implication that does not require knowing workers' beliefs, which makes empirical testing easier. Our empirical analyses are based on a field experiment conducted in an e-commerce warehouse in which we gave employees explicit feedback about the marginal benefit of exerting more effort. As shown in Table 1 in the e-companion to the paper, few studies have estimated the effect of feedback content (see Column 2), and none of these studies has specifically looked at how feedback receivers react to information about marginal benefits (see Column 3). The idea that reactions to feedback are heterogeneous also suggests that feedback may change the distribution of performance within a firm. Such distributional effects have been studied by a number of papers (see Column 4), and our main contribution is to relate these distributional effects to the content of the feedback and to provide quantile estimates, which have been relatively rare in the literature (see Column 5).

3. Analytical Model

Suppose N employees work at a firm and for a given employee i productivity is given by

$$x_i(\rho_i, a_i) = \rho_i a_i$$

where a_i is the employee's work effort and ρ_i is the marginal product of effort. We will assume that this marginal product is not perfectly known to the worker. Specifically, a worker's beliefs about his or her own marginal product will depend on the information that he or she has, which can be of two types. First, the worker may have information based on his or her own observations, which we will denote by y_i . Second, the worker may also receive feedback from the firm, i.e., he or she may have access to information that he or she would not be able to otherwise observe. We will use z_i to denote such feedback.

Suppose employee beliefs are given by a joint density $f_i(\rho_i, y_i, z_i)$, which may be different for each employee. This implies that different employees may have different prior beliefs and may update their beliefs in different ways: specifically, if worker i has observed y_i and has received feedback z_i , his or her

beliefs about ρ_i are given by the conditional density $f_i(\rho_i / y_i, z_i)$. We will assume that the feedback variable z_i is measured in such a way that an increase in its value shifts the conditional distribution of the marginal return according to the monotone likelihood ratio property. In other words, a greater value of z_i means more “positive” feedback, in the sense that when receiving this feedback, the worker believes that a higher marginal return is relatively more likely. One particular case would be that in which the feedback is the worker’s productivity.

Assume that the utility of the worker is given by

$$u_i(a_i; \rho_i) = B(a_i; \rho_i) - c_i(a_i)$$

where B is the gross benefit and c_i is the cost of effort and is increasing in a_i . In this setting, the worker will choose a_i to maximize:

$$\int u_i(a_i; \rho_i) f_i(\rho_i | y_i, z_i) d\rho_i$$

We will use $a_i^*(y_i, z_i)$ to denote the level of effort that maximizes this function. This effort will depend on the information available to the worker, which includes the feedback received.

The main question we are interested in is how worker effort $a_i^*(y_i, z_i)$ will vary with feedback z_i . As a preliminary step, consider the worker’s problem in the absence of feedback. In that case, effort is chosen to maximize expected utility conditional on his or her own information:

$$\int u_i(a_i; \rho_i) f_i(\rho_i | y_i) d\rho_i$$

Let $a_i^*(y_i)$ be such level of effort. According to the model, if the worker received feedback z_i , his or her effort would be given by $a_i^*(y_i, z_i)$; therefore, the effect of feedback is $a_i^*(y_i, z_i) - a_i^*(y_i)$. It is clear that this difference will depend upon the difference in beliefs between the case in which the agent does not receive any feedback and the case in which he or she receives it. Since this difference in beliefs will depend on the value of the feedback variable z_i , we can conclude that $a_i^*(y_i, z_i)$ may in general be greater or lower than $a_i^*(y_i)$. This is consistent with the mixed empirical evidence on the link between feedback and employee performance, as described in the previous section.

To further analyze the effect of feedback, we make more specific assumptions about workers’ utility functions. We allow for two types of worker preferences, one in which workers respond to monetary incentives and another one, which has also been considered in the literature on feedback (e.g. Casas-Arce and Martinez-Jerez 2009; Azmat and Iriberry 2010; Kuhnen and Tymula 2012; Tafkov 2013; Eyring and Narayanan 2018; Gill et al. 2019), in which employees care about their performance relative to coworkers. For expositional purposes, we will first separate the two types of preferences. Beginning with the preferences based on monetary incentives, we assume that the firm pays a fixed hourly wage and an hourly bonus based on the worker’s productivity. In many firms, such as the one we analyze in this paper, it is

common for hourly bonuses to increase with worker productivity. In other words, workers' hourly pay increases with performance. To model this, we assume that the worker's gross benefit is given by:

$$B(a_i; \rho_i) = w(x_i) \equiv (\alpha + \beta x_i)h_i = (\alpha + \beta \rho_i a_i)h_i$$

where h_i is the number of hours worked and the term in parentheses is hourly pay, which includes base pay α and a bonus βx_i that increases with productivity. For simplicity, we assume the number of hours worked is exogenous and normalize it to one ($h_i=1$).

Given these worker preferences, the following result can be derived:

Proposition 1. If $B(a_i; \rho_i)=w(x_i)$, worker effort a_i^ will be increasing in z_i .*

Proof. To prove this result, we use supermodularity theory (Topkis 1978). As shown in Athey (2002) (theorem 1), if $w(x_i)$ is log-supermodular in (ρ_i, a_i) and $f_i(\rho_i / y_i, z_i)$ is log-supermodular in (ρ_i, z_i) , then a_i^* is increasing in z_i . To show that $w(\cdot)$ is log-supermodular, consider two levels of effort a_H and a_L such that $a_H > a_L$. Log-supermodularity is satisfied if the ratio

$$\frac{(\alpha + \beta \rho_i a_H)h_i}{(\alpha + \beta \rho_i a_L)h_i}$$

is nondecreasing in ρ_i for any values of effort. This condition is satisfied in this case, and therefore $w(\cdot)$ is log-supermodular. Moreover, if the density $f_i(\rho_i / y_i, z_i)$ is such that z_i shifts the distribution according to the monotone likelihood ratio, then the density is log-supermodular (Lehmann 1955, Athey 2002). Since the two functions are log-supermodular, a_i^* will be increasing in z_i . QED.

This result implies that while the difference $a_i^*(y_i, z_i) - a_i^*(y_i)$ may be positive or negative, it will unambiguously increase in z_i . This means that more "positive" feedback will have a more positive effect on worker effort than less positive feedback. Note that this result does not require beliefs to follow one particular distribution, e.g., normal. Neither does it require that all workers in the firm have the same prior beliefs: in a firm with many employees, each one may have different beliefs about his or her marginal product and will consequently react differently to the feedback, but the result will still hold for each of them, irrespective of their prior beliefs.

Let us now turn to the case in which the worker's utility depends on his or her productivity relative to the other workers in the firm. Let x_{-i} be the vector of all the workers' productivities except that of worker i , and let $\max\{x_{-i}\}$ be the maximum of these productivities. To model the fact that the worker cares about his or her relative standing in the firm, we will assume that worker i 's utility depends on the difference between his or her productivity and the maximum productivity among coworkers:

$$B(a_i; \theta_i) = v(x_i; x_{-i}) \equiv v_0 + x_i - \max\{x_{-i}\} = v_0 + \rho_i a_i - \max\{x_{-i}\}$$

where $v_0 > 0$ is a constant. The assumption that a worker takes maximum performance as a reference point is consistent with empirical evidence on social comparisons. Indeed, empirical studies have found that

social comparisons tend to be upward rather than downward, i.e., individuals compare themselves to those who are “better” than them: see Gerber et al. 2018 for a meta-analysis of over 60 years of research.⁵

When worker preferences depend on relative productivity, the following result can be found:

Proposition 2. If $B(a_i; \theta_i) = v(x_i; x_{-i})$, worker effort a_i^ will be increasing in z_i .*

Proof. The proof is similar to that of Proposition 1. Since the density is log-supermodular, it suffices to show that $v(\cdot)$ is also log-supermodular. We consider two levels of effort, a_H and a_L , such that $a_H > a_L$, and to show that log-supermodularity is satisfied, we check that the ratio

$$\frac{v_0 + \rho_i a_H - \max \{x_{-i}\}}{v_0 + \rho_i a_L - \max \{x_{-i}\}}$$

is nondecreasing in ρ_i . Since the ratio is increasing in ρ_i , work effort a_i^* will be increasing in z_i . QED.

Although for expositional purposes we have considered the two utility functions separately, we may also assume that the worker’s preferences include both features, i.e., that he or she cares about his or her monetary compensation and about his or her relative standing in terms of productivity. Thus, suppose the gross benefit function is given by $B(a_i; \rho_i) = w(x_i) + v(x_i; x_{-i})$; therefore, the two previous propositions of effort a_i^* will also be increasing in z_i .

In summary, the model provides the following insights. First, feedback may have a positive or negative effect on effort because the effect will depend on how feedback changes workers’ beliefs about their marginal performance. Second, the model underscores the importance of the *content* of the feedback. If employees use the information provided to them to make inferences about the marginal return of their effort, their reactions to feedback will vary according to such information. This will lead to heterogeneous responses to feedback: for instance, if the firm gives all employees information about the marginal benefits of their effort, even though all employees receive the same type of feedback, each one receives a different piece of information, i.e., the “content” of the feedback varies across individuals, and consequently each employee will react differently to the feedback. This implies that an empirical researcher interested in estimating the effect of feedback, e.g., on worker productivity, has to study how the information provided to workers correlates with their post feedback productivity. This requires data not only about which employees received feedback and when but also which specific information was given to them.

⁵ In the feedback literature the papers that use formal models have assumed different reference points, including the maximum, average, or median performance. Casas-Arce and Martinez-Jerez (2019) use a tournament model and therefore the benchmark is the maximum performance. Azmat and Iriberry (2010) assume that utility depends on the difference between own performance and average performance, Chen et al. (2010) assume that it depends on the difference with respect to median performance, and Kolstad (2013) uses a more general model in which the reference point is not restricted to being the maximum, average or median. The model by Kuhnen and Tymula (2012) has only two workers and the reference point is the performance of the other worker.

We think this model provides a useful theoretical framework to understand the heterogeneous effects of feedback, but it relies on several simplifying assumptions that are worth highlighting. We shall focus on three of them that are particularly relevant to the setting in which we conduct our empirical study and to the interpretation of our empirical results. The first one has to do with the information structure of the model. We assume that workers know the effort they choose but have incorrect information about how such effort translates into performance and pay. This is consistent with our empirical setting, where employees choose effort but do not directly observe performance. In other empirical settings this assumption may be unrealistic: for instance, salespersons may have more accurate information about their sales (performance) than about the amount of effort that they dedicated to each client.

The second assumption is related to workers' preferences. We assume that employees care about their relative performance and this preference is "behavioral" in the sense that relative performance enters the utility function directly. However, their reactions to feedback are driven by how they learn from the feedback they receive, i.e., they react to relative performance feedback "rationally" in the sense that they use this information to update their beliefs about the marginal benefit of effort.

The third assumption is the linearity of the bonus with respect to productivity. In our empirical setting the bonus also increases with productivity, but it follows a step function, with discrete jumps at six different productivity levels. These jumps can give rise to interesting effects that are not considered in the model. In particular, workers may choose to exert a particular effort to reach the following bonus step but also to avoid falling back to the previous step. This would have implications for their reactions to feedback, which we discuss in the empirical part of our study. Specifically, an employee who is further away from the following bonus step would have a lower incentive to increase effort because the cost of reaching the higher bonus is greater, but being further away from the following step implies being closer to the previous step, which may create an incentive to avoid falling back to the lower bonus level.

4. Institutional setting

We conducted the experiment at the warehouse distribution center of a Fortune 500 company located in the US midwest.⁶ We refer to the site as Midwestern Warehouse (MWW). MWW had over 1,000,000 square feet of space and was responsible for handling 80,000 different stock-keeping units (SKUs). There were approximately 130 hourly employees who worked at the site over three shifts. On a daily basis, MWW shipped an average of 8,000 orders comprising approximately 100,000 total items. The facility was divided into four sections that were physically separated from one another. The sections handled different SKUs and were separated for product grouping purposes, but the tasks of the workers were identical in all sections.

⁶The company has many distribution in the US and throughout the world, but we were given access only to this one.

Each employee was hired for a specific section when the section needed a worker and remained in the same section thereafter. Each section was divided into approximately 40 work areas referred to as stations. The technology used at MWW was industry standard, i.e., what is typically present in warehouses that pick, pack, and ship products to customers. Facilities of this type have become increasingly common with the growth of e-commerce, as firms stock products centrally and then ship them across a region, country, or globally. According to the Bureau of Labor Statistics, this industry employs over 1.8 million people in the United States (Bureau of Labor Statistics, 2022).

Once an order came into the warehouse, a specialized Warehouse Management System (WMS) routed plastic tote bins via a conveyer belt to the appropriate stations so that employees could select the items needed for the order. Employees would move to the stations where there were bins to fill the orders from those stations and would know which stations to go to since that information was displayed on large screens. At each station, the employee would scan the bin and an electronic board would indicate the items that he or she should pick from the station and place in the bin. The employee would then walk within the station, select the correct items and quantities, scan them, and place them in the plastic bin. The employee carried out the entire order from the station. After all the items from the station had been selected, the bin would then move to either another station or to the packing area via the conveyer belt, where items were placed into a box, and the box was labeled for shipping. In the picking stations of the facility, the workload of each worker was independent of the work done by others, and there were no spillover effects from the work of one worker to another.

The WMS also collected performance data. When the system was installed, engineers calculated how long it should take to prepare each order, taking into account the number of items required and their location. These standard times were used thereafter to calculate individual performance: every day, the WMS collected data on the total time each worker actually took to prepare each order, and the ratio of standard time to actual time was then used as a measure of performance.⁷ Typically, a job task lasted between 20 seconds and 3 minutes, depending on the number of items in the order and the location of the items within the station. The WMS calculated a rolling performance metric for every employee.

Weekly pay was equal to hourly pay times the number of hours worked during the week, and hourly pay was equal to base pay plus an hourly bonus based on weekly performance. The hourly bonus increased with performance, as shown in Figure 1, and incentive pay could account for over 30 percent of compensation. At the beginning of the week, employees received a payroll slip that indicated how much

⁷ For instance, if a bin arrived at Station 25 and the order called for three units of Item A and two units of Item B, the WMS would know that the order should take 50 seconds to complete. If an employee completed the task in 60 seconds, the performance score would be 83.3 percent (50/60). Similarly, if an employee took 43 seconds to complete the task, the performance was 116.2 percent (50/43).

would be deposited into their account for the hours worked the previous Monday through Sunday. The payroll slip displayed the gross pay, taxes and other deductions, net pay and total hours worked, but did not contain explicit information about their performance or about the split between fixed and variable pay.

During the shift, employees could find their performance displayed at special monitors throughout the warehouse. Monitors displayed the employee's name, current station, and individual performance on the shift in real-time. However, it was costly for employees to correctly estimate the benefit of working harder for two reasons. First, some information was not available to them. This was the case for relative performance, which was not displayed on the monitors. Second, the information shown on the screens was offered in real time but only for that day. At the end of the shift, that information was changed for the people in the next shift. If an employee wanted to use the available information to estimate the benefits of his or her effort, substantial costs would be involved: he or she would have to record this information every day at the end of the shift, compute the weekly weighted average of performance, introduce this information into the nonlinear bonus formula given by the company, and analyze the potential benefits of working harder.

Most workers entered the facility as temporary workers (hired through a temporary help agency), and after a three-month probation period became permanent employees. Workers could also enter directly as permanent employees, but this required a longer waiting period due to administrative paperwork. MWW defined poor-performing employees as those who consistently performed below 70 percent relative to the standard time throughout the week. Poor-performing employees had to go through training for a week. If they continued to perform poorly, they were offered coaching and job shadowing for a second week⁸, and if that did not result in improved performance, they were dismissed.

5. Experimental design

At the beginning of each shift on Mondays, managers met with all the employees in a given section. Two weeks prior to the experiment, an overview of the project was presented to all the employees of the shift and section during that meeting. The site manager introduced one of the researchers, who explained that the team was independent of senior management and that it comprised business school professors who wanted to better understand the facility's operations.⁹ The researcher informed them that each employee would receive a sheet of paper every Monday at their weekly meeting. He asked that all employees read

⁸ We also examined whether our results held when controlling for the fact that some workers whose performance was under 70 percent might have received more feedback than the rest. Our results remained unchanged.

⁹ A meeting was conducted for each of the four sections, since these sections were separated and distant from each other. This was then repeated for each of the 3 daily shifts.

this piece of paper before beginning their work for the week. We used that meeting to ask them to sign a form giving us consent to receive the sheet of paper every Monday (as required by the IRB).

We also made it very clear to them that the standards to which they were subject would not be changed and that there would be no reprogramming of the way work was done in the unit. We did this with the aim of avoiding the so-called ratchet effect (Weitzman 1980; Freixas et al. 1985; Leone and Rock 2002), i.e., quota restriction behaviors that occur when workers are afraid of being subjected to tougher goals if they improve productivity. This also means that relative concerns, if found, were less likely to be caused by the fear of being ostracized by other workers. In this warehouse, workers' tasks were independent from one another, the layout was fixed by management and the process was highly systematized. Thus, there was a small risk that workers would take actions to undermine others' performance (e.g., by leaving things in the wrong place or giving other workers wrong instructions) in order to improve their own relative standing.

We applied two feedback treatments -individual and relative feedback- that were designed to appeal to each type of preferences considered in the model and we also a third feedback treatment—"full feedback"—which provided all pieces of information at the same time. We employed a 2x2 design, assigning these three treatments of feedback and the control across the four different sections of the factory. Each of the four sections was initially randomly assigned to one of the four different options: individual feedback, relative feedback, full feedback, or control. After two weeks, each group was then rotated. The experiment was run for a total of eight weeks, and every section of the facility spent two weeks in each of the four different options. Thus, all the sections in our unit were subject to all the conditions at some point during the exercise.

For the individual feedback treatment, we told employees how close they were to the next level of performance (i.e., the level that would entitle them to earn the next higher bonus)¹⁰ and how much additional income they would have earned at the next bonus level. We also informed them about their previous week's performance. For the relative feedback treatment, employees were told about their own level of performance, their performance ranking relative to everyone working in their section of the factory (i.e., the percentage of employees with lower performance), and the level of performance of the top performer. The full feedback treatment included both the individual and the relative feedback messages. Giving all the information together allowed us to explore the effect of each type of feedback while controlling for the other type and thus to control for the extent to which the effects of giving individual (relative) information could be driven in part by guesses about their relative standing (monetary gains). In all cases, the measure of performance that we used was weekly performance as defined by the firm. Employees in the control

¹⁰ The information we gave was framed as potential gains. Alternatively, we could have framed it highlighting both potential gains and losses, but because each frame could lead to different behaviors (Kahneman and Tversky 1979) this could add noise to the question. Therefore, for simplicity and to stick to a more realistic setting (highlighting gains is the message the company gave workers when hiring them) we chose to focus on gains.

group received a standard message about safety that was written throughout the facility and always discussed at every weekly meeting. The standard templates that were used to create the messages given to each individual employee, as well as the description of the order of the treatments are shown in the e-companion of the paper. On Sunday nights, we downloaded the data from the WMS, made the calculations to produce feedback to deliver, and printed the message on a piece of paper. This was then distributed to each employee on Monday before they began their shift. After the 8-week experimental period, workers stopped receiving feedback through weekly pieces of paper.

The layout of the four sections reduced the risk of contamination across the different treatments because workers assigned to a section did not see or interact with workers in other sections while working. Each of the four sections had different parking lots, lunch locations, break areas and bathroom areas, as well as different lunch break schedules. In our field observation and qualitative interviews with the management of the warehouse, we realized that nobody was going from one area to another to see friends. Before running the experiment, we tested for differences among the four sections, and although we found no significant difference in terms of average performance, we did find some differences regarding other relevant characteristics (please, see the e-companion for a full description of these tests).

Rotating the treatments across sections allowed us to observe every individual under the different treatments, which given the evidence of some heterogeneity across sections would help us ensure that results are not driven by differences between sections. This enabled us to control for extraneous unobserved characteristics of participants and sections that could be influencing the results (Charness et al. 2012; List et al. 2011). Furthermore, given the relatively small number of employees per section, the within-individual design allowed us to have more statistical power as more people were exposed to each treatment. However, the experiment could lead to the wrong conclusions if there were reasons to suspect that the order of the treatments could affect the results (i.e., if there is no independence of the multiple exposures). To address these concerns, we designed the rotation of the treatments to ensure that different individuals received different types of feedback in different orders (as shown in e-companion of the paper).¹¹ We also explore the potential existence of order effects in the analyses section.

6. Data and Variables

We focus on permanent workers to obtain a more homogenous sample since these workers have already passed the screening process. Permanent workers also provide a more comparable sample since many of

¹¹ We could not design our experiment to control for all possible order effects without avoiding potential contamination, as this would have required at least 16 different sections.

the temporary workers did not intend to stay there for a long time.¹² The number of permanent workers varied per section (see Table 3A for the number of permanent employees in the week immediately before the experiment)¹³. The final dataset includes 1,105 employee-week observations. In addition to the eight weeks of the experiment, we collected data in the six consecutive weeks that immediately preceded the experiment and in the eight consecutive weeks that immediately followed it. (See e-companion for a full description of the variables per section as well as for the final composition of the sample).

The left-hand-side variable in all the regressions is the logarithm of individual performance, as defined by the company. In Figure 1, we plotted a histogram of performance (for the six weeks prior to the experiment) and the step function that represents the incentive scheme, i.e., the hourly bonus that an employee is entitled to as a function of performance. Performance is concentrated in the interval for which the bonus function is increasing, which suggests that earning the bonus was neither too “easy” (in which case most observations would cluster around the flat portion of the bonus scheme corresponding to above 130 percent performance) nor too “difficult” (in which case performance would cluster below 80 percent).

[Insert Figure 1 about here]

Since we are interested in the performance effect of feedback content, the main right-hand-side variables are measures of the information given to the workers. To operationalize the feedback information given to each employee, we construct several variables, hereby referred to as “information variables,” that take the value of the specific information provided and zero if that information was not provided. We use two alternative approaches to define the information variables.

In the first approach, we define four information variables that do not treat information differently depending on whether it was given independently or together with the other type of feedback. For individual feedback, we construct two information variables: “distance to next level” and “bonus increase”, which take the value of the specific information given regarding distance and bonus when individual feedback was given and zero when it was not given. We use these two variables as proxies for the marginal monetary benefit of effort. A greater value of Distance to next level indicates that a given amount of additional effort will yield less benefits, since the next level of bonus is more distant. Therefore, an increase in this variable is a proxy for a reduction in the marginal benefit of effort.¹⁴ For Bonus increase, a greater value of this

¹² Approximately 80 percent of temporary workers quit after two weeks and many quit after one week. Managers told us it was expected that 15–20 percent of temporary workers would not return for the second day of work.

¹³ The number of permanent workers per session was not constant throughout the whole period but there was very little variation. Workers did not move across sections during our study period.

¹⁴ Since the bonus scheme is a step function, a greater distance to the next level also implies a smaller distance to the previous level. If employees understand that being close to the next level also implies being far from the previous level, an increase in Distance to the next level would have two effects: a reduction in effort because it is more difficult to reach the next level of bonus and an increase in effort because it is easier to fall to the previous level of bonus. If these two effects are taking place, then our empirical analyses underestimate the effect that feedback would

variable indicates that a given increase in work effort yields a greater monetary payoff. Hence, an increase in this variable is a proxy for an increase in the marginal benefit of effort. Consequently, a lower Distance to next level and a higher Bonus increase should be associated with greater effort.

Similarly, we construct two information variables for relative feedback that take the value of the relative feedback when it was given and zero when it was not given. These variables are “Performance quantile” (percentage of employees in his or her group that the focal employee has outperformed) and “Distance to maximum performance” (difference between the performance of the best performer and the focal employee’s performance). Given the effort that the employee has made in the previous week, a higher Performance quantile should lead the worker to believe that their marginal product of effort is higher than it would be if this variable were lower. Likewise, a smaller Distance to maximum performance must lead to the belief that the marginal product is higher than it would be if the distance was greater. Consequently, a higher Performance quantile and a lower Distance to maximum performance should be associated with greater effort.

As an alternative, we use a second approach, which is to construct the information variables in a way that allows the information to have a different effect depending on whether it was given separately or together with the other type of feedback. For example, for individual feedback, we create two “distance to next level” variables, one that takes the value of the distance information given if it was communicated independently and zero otherwise (Distance to next level (IF)) and another that takes the value of the distance information given if it was communicated together with relative feedback and zero otherwise (Distance to next level (FF)). We apply the same procedure for the other information variables and obtained 8 alternative information variables (Distance to next level (IF), Bonus increase (IF), Performance quartile (RF), Distance to max performance (RF), Distance to next level (FF), Bonus increase (FF), Performance quartile (FF), and Distance to max performance (FF)).

Finally, we include a set of control variables. First, we control for the performance of individuals before and after the experiment week, including two indicator variables: *pre-experiment week* and *post-experiment week*, which take the value of 1 if the individual-week observation corresponds to the weeks before and after the experiment, respectively, and zero otherwise. Second, since the information variables take the value of zero in two different cases –(1) when the information is not given or (2) when the information is given but the marginal benefit of effort is zero– we include two additional control variables: Individual Performance Max, which takes the value of 1 if they had reached the maximum level of bonus in the previous week, and zero otherwise, and Relative Performance Max, which takes the value of 1 if the

have in the case of a linear bonus scheme. In our empirical analyses we consistently find a negative effect of Distance to next level on performance, indicating that the effort reduction effect exceeds the “fall-down” effect.

individual was the top performer in her unit in the previous week and zero otherwise.¹⁵ See the e-companion for a detailed description of how the main empirical variables relate to the theoretical variables as well as the definition of the variables used in the main empirical analyses.

We estimate all our regressions using employee fixed effects to control for any unobserved fixed worker characteristics. We cluster the standard errors by worker to take autocorrelation into account. Table 1 shows summary statistics for the main dependent variable and the information variables.

[Insert Table 1 about here]

7. Results

7.1. Feedback effects on performance in the workplace: descriptive results

We first plot the kernel distribution of performance for the workers before the treatment (the 6 weeks prior to the treatment) and during the 8 weeks of the treatment: see Figure 2. shows four graphs. Figure 2a shows the distribution of performance for all the sections before the treatment, as well as the performance for all the sections that were affected by a feedback treatment during the weeks of the treatment (i.e., excluding the performance of the sections that received only the safety message). We observe that the performance distribution does not seem to move to the right or left, but that the treatments seem to compress the performance distribution. Figures 2b to 2d show the performance distribution for each treatment separately. For comparison purposes, we plot only performance for the first two weeks of each treatment, which, because of the way we conducted our experiment, corresponds to performance in only one section, together with the performance distribution in that particular section pretreatment. We do not observe an overall effect on the average performance, but we see that the treatments affect the shape of the distribution; particularly when we provide individual feedback and relative feedback separately, the distribution seems to become more concentrated.

[Insert Figure 2 about here]

We also explore whether there is an average treatment effect of each of the three treatment conditions (IF, RF and FF) in a more systematic way. To do this, we use OLS regression with worker fixed effects including three dichotomous variables, one for each treatment. Estimates, displayed in Model 1 of Table 2, show that none of the treatment variables has a significant effect on performance. This is also consistent with what we observed in the performance distribution figures and in the descriptive statistics: the feedback policy does not seem to matter for changing average performance. We next turn to explore our main hypothesized effects, those regarding the provision of information about the marginal benefits of effort.

¹⁵ After deleting the temporary workers in our sample, there were no individuals for whom the information variable Performance quantile took a value of zero.

(Additional analyses comparing the average performance effects of the treatments are presented in the e-companion of the paper).

7.2. Content feedback effects on average performance in the workplace: linear estimation results

Table 2 shows the results of our model estimations, using the four information variables (Model 2, Table 2), or allowing the effect to be different in each case, using the eight information variables (Model 3, Table 2). The individual feedback information variables show a significant effect on performance. Specifically, the performance distance to the next level has a negative effect in both Models 2 and 3 (albeit only when given together with relative feedback). Using the coefficient estimated in Model 3 (Distance to next level FF), we find that informing individuals that they are 9 points away from the next level versus informing them that they are 1 point away from the next level leads to a reduction in performance of 1.54 percent. The bonus difference, on the other hand, has a positive effect on performance in both specifications, albeit only at a 10 percent level in Model 3. Using the estimates from Model 3 (Bonus increase IF), we find that telling individuals that they could make \$0.75 more in bonus (the maximum increase) versus \$0.25 more in bonus (the minimum increase) leads to a performance increase of approximately 1 percent.

Both effects have the predicted sign. First, if distance to the next level is greater, this means that the marginal benefit of effort is lower: for employees who are very close to the next level, a small increase in performance suffices to increase the bonus, but for employees who are far from the next level, a small increase in performance may not be enough to earn a larger bonus, so that the marginal benefit would be zero. Since a greater distance implies a lower marginal benefit, we expect a negative effect of distance on worker performance. Second, if the bonus difference is greater, then the marginal benefit of effort is greater, which implies that the bonus difference should have a positive effect on individual performance. With regard to the relative feedback information, the effects also have the predicted sign, but the levels of significance are lower: we find that the performance quantile has a significant, positive effect in Model 2 and in Model 3 (albeit only in the treatment group that receives both types of feedback), but Distance to maximum performance does not have a significant effect in either model. Finding that the performance quantile has a significant effect under the full feedback condition suggests that individuals have behavioral preferences for a higher-standing position. Using the estimates from Model 3 (Performance quantile RF), we find that telling individuals that their performance was in the 75th percentile in their section versus the 25th percentile results in a performance increase of approximately 0.88 percent.

Another relevant finding from Model 3 of Table 2 is that the coefficients of the information variables in the individual or relative feedback cases are not significantly different from their full feedback counterparts. This suggests that employees do not use the information we give them relative to the marginal benefit in terms of monetary gains to make guesses about the marginal benefit in terms of relative standing gains and *vice versa*. Regarding the controls, the pre-experiment period shows a positive and significant

coefficient and the post-experiment period shows a negative and significant coefficient. To interpret these estimates, we must keep in mind that the omitted dummy is for being in the control group, which means not receiving any feedback while other workers are receiving some. We find that absence of feedback is associated with lower performance when employees who do not receive feedback know that others are receiving feedback or when feedback was given to them at some point but is no longer given.

Because we are conducting hypothesis testing for multiple treatment groups, we need to account for the fact that we may reject more null hypotheses than we should (Floyd and List 2016; Young 2019). To address this potential concern, we follow previous work (Casas-Arce et al. 2017) and first test the joint significance of the three treatment coefficients using an F test. The results are reported at the bottom of Table 2 and reveal that the main treatment variables are jointly statistically significant at a 10 percent level in Model 2, but not in Model 3. We also use randomization testing to test the null hypothesis of no treatment effects: using 1,000 random draws from the treatment vector, we use the distribution of the coefficients of the randomized regressions to test the statistical significance of the estimated treatment coefficients. The p values computed under this randomization testing confirm the conclusions reached when looking at the p values that result without randomization testing. Finally, we also test for joint significance of all treatments, computing a new p value for the Wald statistic of the original regression using the distribution of the Wald statistics of joint significance of each of the regressions with the simulated treatment coefficients. We again find that the treatment information variables of Model 2 are jointly significant at the 10 percent level.

7.3. Content feedback effects addressing potential order effects

As mentioned before, the design of the experiment could lead to the wrong conclusions if order effects are present: for example, if individuals react to the information variables that provide individual feedback differently depending on whether they had received relative feedback versus the control information message before. We conduct a series of additional regression analyses to explore potential order effects.

First, we perform an analysis using only information from the first two weeks of the experimental period. This is equivalent to a between-individual design, as during this first two-week period individuals in each section were only receiving one type of treatment (individual, relative, full, or control) without a different preceding treatment. Thus, order effects should not be an issue in this case. Looking only at the first two weeks also controls for potential “experimenter demands” that could happen if participants realized that the goal of the study was the feedback system and manipulated their performance to be best during the feedback system they preferred. These results are shown in Table 3 (Models 1 and 2) and reveal some similarities and some differences with respect to the analyses that use the sample from all the experimental weeks. In terms of similarities, Bonus increase and Performance quantile show a positive and significant coefficient. In terms of differences, the variables related to distance do not show a negative or significant effect, and indeed, Distance to maximum performance shows a positive and significant effect. Section 2

(the section that received relative information feedback independently in the first two weeks) shows a significantly smaller average distance with respect to the maximum performance in the weeks prior to the experiment compared to the rest of the sections (41.12 versus 54.61, $p=0.000$).¹⁶ Thus, one reason for this positive coefficient could be that for relatively smaller distances, individuals are encouraged to exert more effort when they learn their distance to the maximum performance.

Next, we control in the models for the order in which the different pieces of information were given. For each information message given (i.e., IF, RF, FF, Control), we construct a count variable that takes the value of 1 if the information was given in the first two weeks, 2 if it was given in weeks 3 and 4, 3 if it was given in weeks 5 and 6, and 4 if it was given in weeks 7 and 8. These four ordinal variables (Order IF, Order RF, Order FF, Order control) are intended to capture an aspect of learning effects: if individuals have received other information before, e.g., those who received IF after having received all other types of treatments, may learn less about their marginal benefit of effort and therefore react less to the new information provided. Thus, if order effects are present, we should expect a negative coefficient of these order variables when including them in the model. When we introduce these four variables as controls in the model, we find some evidence of order effects (Models 3 and 4, Table 3). Specifically, the coefficient of Order FF is negative and significant in both models, which suggests that the effect of the feedback given under the FF condition is larger the first time individuals receive the feedback (the effect of the FF information variables decreases when feedback is provided in later weeks). The results regarding the information variables continue to be consistent with the model's predictions, although we lose significance in some of the coefficients.

[Insert Table 3 about here]

As a final test, we also estimate a model allowing for the effects of each of the pieces of information feedback, including the control information, to vary depending on the week period in which the information was provided (analyses available upon request). Thus, for each information variable we construct four variables. For example, for Bonus Increase (IF) the four variables are called Bonus Increase (IF) weeks 1&2, Bonus Increase (IF) weeks 3&4, Bonus Increase (IF) weeks 5&6 and Bonus Increase (IF) weeks 7&8 and are defined as follows. Bonus Increase (IF) weeks t &($t+1$) (for $t=1, 3, 5, 7$) takes the value of the information given regarding the bonus increase but only if the information was given in weeks t and $t+1$ of the experiment, and 0 otherwise. We apply the same procedure for each information content given: Distance to next level (IF), Performance Quantile (RF), Distance to maximum performance (RF), Bonus Increase

¹⁶ Average distance was also smaller in Section 3, which received relative and individual feedback together, but the difference was not statistically significant.

(FF), Distance to next level (FF), Performance Quantile (RF), Distance to Maximum Performance (RF), and the control information. This leads to a total of 36 information variables. We introduce these new information variables in the model (instead of the information variables of the main analyses in Table 2) and perform Wald tests to check for differences in the coefficients of each information variable across the four periods (e.g., differences across the coefficients for Bonus Increase (IF) weeks 1&2, Bonus Increase (IF) weeks 3&4, Bonus Increase (IF) weeks 5&6, and Bonus Increase (IF) weeks 7&8). In the presence of order effects, we should see statistical differences among the coefficients. We found only two instances in which there were significant differences across weeks: Distance to next level (IF) and Distance to next level (FF). In conclusion, although we find some evidence of order effects, our results are in general robust to controlling for the order effects.

7.4. Feedback effects on the distribution of performance in the workplace: quantile regressions

We next turn to investigate the distributional effects of feedback by estimating quantile regressions to explore which part of the distribution is more affected by the treatments. For simplification purposes, in what follows we focus on the Model 2 specification.¹⁷ As suggested in Table 2, individual feedback may increase or reduce performance, depending on employees' learning about the marginal benefit of effort. Such effects might lead to changes in the distribution of performance in the workplace, which may become more dispersed or more concentrated around the median. The distribution of performance shown in Figure 2 suggests that such distributional effects took place, at least for some of the treatments.

Distributional effects can take place because differences in performance can be partly due to differences in employees' beliefs about the marginal benefit of effort. Hence, if feedback changes workers' beliefs, it will also change the distribution of performance. As an example of why feedback may have distributional effects, suppose that all employees have the same cost function for effort, suppose the function is convex, and assume differences in performance are solely due to beliefs, i.e., low performers are employees who believe that the marginal benefit of effort is low and high performers are those who believe that it is high. If all receive the same piece of information on the marginal benefit, the effect will be different in each group: low performers will learn that the marginal benefit is greater than they thought and will increase their effort, while high performers will determine if they were overestimating the marginal benefit and will reduce their effort. We consequently expect a similar piece of information to have a different effect on performance in different parts of the distribution.

To test for distributional effects, we estimate quantile regressions using the "quantile regression for panel data" model proposed by Powell (forthcoming). This approach allows us to estimate treatment effects on the quantiles allowing for unobserved employee fixed effects and, to our knowledge, has not been used

¹⁷ We performed all the other analyses of these paper using also the specification presented in Model 3 and reached the same conclusions presented here.

in the earlier literature on feedback.¹⁸ The results, shown in Table 4 (Models 1 to 3), indicate that almost all the information variables have significant effects on some of the quantiles, which is consistent with the idea that employees take into account all the information disclosed. In addition, the signs of the effects that are statistically significant are the same for the different quantiles. The models also show that the effects tend to be stronger and more significant for the 25th and 50th quantiles, suggesting that top performers tend to react less to feedback information.¹⁹

[Insert Table 4 about here]

7.5. Additional analyses

We perform a series of additional analyses (shown in the e-companion) to provide suggestive evidence of our main mechanism as well as to explore the robustness of our results to alternative specifications. Specifically, we show that the relationship between prior performance and performance became weaker during the experiment period, which is consistent with the hypothesized learning mechanism. We also rule out that our results are only driven by a Hawthorne effect, as we continue to find similar effects when we remove data from the last two weeks immediately before the experiment started, when participants learned that the study was going to take place and could have already reacted to the experimenter introduction.

In addition, we find that our results are robust to alternative specifications, including the use of an alternative operationalization of the information variables (i.e., taking into account whether the worker was in the individual performance incentive zone or not), excluding the post-experiment performance, allowing for two-way clustering (worker and section) of standard errors, and removing data from workers who joined or left our sample during the experiment period.

8. Discussion

On the whole, our analyses give strong support to the theoretical predictions. With regard to *individual feedback*, employees receive information about the extra bonus they would earn if they improved their performance and when this bonus is greater, the effect on performance is also greater. Employees also receive information about how far they are from earning a greater bonus, and when this distance is greater, the effect on performance is smaller. We find these effects in both linear and quantile regressions.

¹⁸ An alternative fixed-effect estimation approach is Canay's (2011) two-step procedure. The main advantage of Powell (forthcoming) is that it does not constrain unobserved heterogeneity to be additive and therefore allows such heterogeneity to have different effects on different parts of the distribution.

¹⁹ To further explore whether some of the effects could be driven by extremely low performers who are very far from receiving a bonus, we remove the lowest performing employees and limit the sample to employees with performance above 70 percent. When we do so we continue to find a negative coefficient for the distance variable, although it is no longer significant, suggesting that the effect of distance is stronger for these low performers.

Regarding *relative feedback*, we also find that the effects are consistent with the theory but in some cases the effects are not statistically significant. In the experiment, employees learned their performance quantile and how far they were from becoming the top performer. In the linear regressions, the performance quantile has a significant and positive effect only in one of the models, and the distance variable is almost always insignificant. In the quantile regressions, the effects of relative information are significant for the bottom and the medium performers but are not significant for the top performers.

Our study has important implications for both theory and practice. Theoretically, we provide a model of how employees use feedback to update their beliefs and reoptimize their choice of effort. The model provides testable implications of how feedback content should influence performance. Such predictions do not require knowledge of employees' prior beliefs and do not rely on assuming that beliefs follow a specific statistical distribution.

From a practical standpoint, our study has important implications for the design of information systems in organizations. Our findings suggest that the same feedback policy will have different effects depending on the content of the feedback. For instance, if the policy is to tell employees information about their previous week performance, the effect will differ cross-sectionally (across employees) and longitudinally because performance varies both cross-sectionally and longitudinally. Our study also suggests that a feedback policy may have very different effects depending on employees' beliefs about how beneficial it is to exert an additional unit of effort.

Our results are of particular relevance to operations managers who need to provide feedback in a warehouse setting, which is different from a typical assembly line process. In an assembly line, employees have defined tasks and roles and these tasks rarely change, employees can be taught how to perform all the tasks for their station, and evaluating performance can be straightforward. However, warehouse operations, and the autonomy they provide the employee to complete the task, are inherently different from a typical assembly line facility (Sun et al. 2021): the task itself must be accomplished, but how each employee goes about accomplishing it is not pre-defined. This adds complexity to the training, compliance, and improvement process, especially when employees are physically dispersed (Staats et al 2016).

Our study suggests that feedback changes employee performance as long as employees learn from feedback, and this has practical consequences for firms providing feedback on a regular basis. When feedback is frequent and the work environment is relatively stable, employees may eventually reach a level of knowledge of the relevant parameters (in our case, the marginal benefit of effort) such that further feedback will have little impact on performance. When feedback is less frequent or the work environment is more complex or changing, the information that employees receive with the feedback will have a greater learning effect and a greater effect on performance is to be expected.

Our study is also suggestive of the complementarities that exist between operations management, information sharing, and incentive design. The technology that the firm uses for the efficiency of its operations generates the data that allow it to implement its incentive system and to provide detailed feedback. Although the establishment uses state-of-the-art technology to manage its operations and the incentive system is also quite sophisticated, information sharing matters, suggesting that productivity is not fully driven by technology or by the incentive system.

Given the effects of feedback on productivity, an interesting question is whether these effects translate into profits. Employees who learn that their productivities are high may seek outside job opportunities, which may lead to turnover or to wage renegotiation and a subsequent increase in costs. If the firm faces competition from a dominant rival who is able to pay higher wages, giving employees feedback may have negative consequences on profits. We would expect this negative effect to be larger when the feedback is relative, since information on relative productivity is more valuable for employees to successfully move to a better paying job. When feedback is individual, employees learn about their productivity but they do not know how many co-workers have similar productivities, which makes it harder for them to evaluate their outside job opportunities. This suggests that firms with less market power would rely more on individual feedback than firms with more market power. In this study we have not analyzed the link between market power and the type of feedback but we think that this is an interesting question for future research.

Another question that is worthwhile addressing is whether there are benefits to disclosing feedback on employee performance to third parties like clients or suppliers. If a client firm has access to feedback about the employee who is mostly dealing with it, it may use the information to make inferences about the value of the service that the employee provides. Such learning may have greater consequences when the feedback is relative than when it is individual. Under relative feedback, if a client learns that the employee it deals with is a relatively low performer, the client may infer that the firm is not doing enough to provide a good service. However, when the feedback is individual making this type of inference will be more difficult, since the client may not have an accurate idea as to whether a given score can be considered high or low. There may also be differences across clients, as more experienced ones may have the capacity to extract information on relative performance out of individual performance feedback. We think that these questions are indicative of the potential for future research of our learning-based model.

Our study has a number of limitations. First, to avoid contamination issues we decided to apply the same treatment to all the employees within the same section and we used a pre-defined order of treatments. This limited our ability to explore potential sequencing effects, i.e. effects of the order in which employees receive feedback. Second, we found that individual and relative performance feedback changed the distribution of performance but we were limited in our ability to investigate the mechanisms that drive these effects. We observed that the performance of the bottom performers improved after the receipt of individual

and relative feedback but we do not know whether these effects were due to greater effort being exerted or effort being exerted in a smarter way.

Another important question is whether our findings are generalizable to other settings, in other words, the external validity of our experiment. In our setting tasks were simple and productivity was easy to measure. We believe that our results may apply to settings with similar characteristics, in particular, to those in which time is a key performance indicator. We think that when performance is difficult to measure and includes a subjective judgment from the evaluator, such as with more complex tasks, the provision of individual and relative performance feedback may actually have a greater impact, since workers' prior beliefs are likely to be less precise.

Another feature of our experiment is that employees received nonlinear incentive pay. In this context feedback about the marginal benefit of exerting effort may be particularly valuable, but in cases where there is a linear-pay formula, workers may already have a better sense of the marginal benefits of exerting effort. Moreover, we were not able to change the incentive structure offered to employees. Being able to manipulate both feedback and the type of incentives at the same time would be useful to better understand the boundary conditions of our findings. Moreover, how feedback is provided may also affect the results. In our context, feedback was given in a private way, so the effects of relative feedback should be due to self-esteem rather than to other relative concern mechanisms, and in contexts in which information about relative standing is given in a public way we may expect even stronger effects of giving relative feedback (see, for example, Hannan et al. 2008; Tafkov 2013).

Acknowledgements

We thank Anil Arya, the Senior Editor and two anonymous reviewers for their comments and suggestions. Earlier versions of this paper were presented at the Strategic Management Society annual meeting (Berlin, September 2016), Madrid Work and Organizations Workshop (May 2017) and Wharton People and Organizations conference (Philadelphia, PA, October 2017), and at seminars at CUNEF and Indiana University. We are grateful to participants at these conferences and to Benjamin Barber for their comments and suggestions. All remaining errors are our own.

Jaime Ortega acknowledges partial financial support from grant PGC2018-098767-B-C21 (MCI/AEI/FEDER, UE) and from the Madrid Government (Comunidad de Madrid-Spain) under the Multiannual Agreement with UC3M in the line of Excellence of University Professors (EPUC3M12), and in the context of the V PRICIT (Regional Programme of Research and Technological Innovation).

Rocio Bonet acknowledges partial financial support from the Spanish Ministry of Economy and Competitiveness (MCIU), State Research Agency (AEI) and European Regional Development Fund (ERDF) Grant No. PGC2018-098767-B-C22.

References

- Aral, S., E. Brynjolfsson, L. Wu. 2012. Three-way complementarities: Performance pay, human resource analytics, and information technology. *Management Science*, 58 (5), 913-931.
- Athey, S. 2002. Monotone comparative statics under uncertainty. *Quarterly Journal of Economics*, 117 (1), 187-223.
- Azmat, G., N. Iriberry. 2010. The importance of relative performance feedback information: evidence from a natural experiment using high school students. *Journal of Public Economics*, 94 (7-8), 435-452.
- Azmat, G., M. Bagues, A. Cabrales, N. Iriberry. 2019. What you don't know... can't hurt you? A natural field experiment on relative performance feedback in higher education. *Management Science*, 65 (8), 3714-3736.
- Baldassarre, L., B. Finken. 2015. GE's real-time performance development. *Harvard Business Review*, 93 (7/8).
- Banker, R. D., G. Potter, D. Srinivasan. 2000. An empirical investigation of an incentive plan that includes nonfinancial performance measures. *Accounting Review*, 75 (1), 65-92.
- Blanes i Vidal, J., M. Nossol. 2011. Tournaments without prizes: evidence from personnel records. *Management Science*, 57 (10), 1721-1736.
- Buckingham, M., A. Goodall. 2015. Reinventing performance management. *Harvard Business Review*, 93 (4), 40-50.
- Bureau of Labor Statistics, 2022. About the warehousing and storage subsector. Accessed on July 1, 2022: <https://www.bls.gov/iag/tgs/iag493.htm>
- Canay, I. A. 2011. A simple approach to quantile regression for panel data. *Econometrics Journal*, 14 (3), 368-386.
- Cappelli, P., A. Tavis. 2016. The performance management revolution. *Harvard Business Review*, 94 (10), 58-67.
- Cappelli, P., A. Tavis. 2018. HR goes agile. *Harvard Business Review*, 96 (2), 46-52.
- Casas-Arce, P., C. Deller, F. Martínez-Jerez, J. M. Narciso. Knowing that you know: incentive effects of relative performance disclosure. *Review of Accounting Studies*. Forthcoming.
- Casas-Arce, P., S. M. Lourenço, F. A. Martínez-Jerez. 2017. The performance effect of feedback frequency and detail: Evidence from a field experiment in customer satisfaction. *Journal of Accounting Research*, 55 (5), 1051-1088.
- Casas-Arce, P., F. A. Martínez-Jerez. 2009. Relative performance compensation, contests, and dynamic incentives. *Management Science*, 55 (8), 1306-1320.
- Casas-Arce, P., F. A. Martínez-Jerez, V. G. Narayanan. 2017. The impact of forward-looking metrics on employee decision-making: The case of customer lifetime value. *Accounting Review*, 92 (3), 31-56.
- Charness, G., U. Gneezy, M. A. Kuhn. 2012. Experimental methods: between-subject and within-subject design. *Journal of Economic Behavior and Organization*, 81 (1), 1-8.
- Chen, Y., F. M. Harper, J. Konstan, S. X. Li. 2010. Social comparisons and contributions to online communities: a field experiment on movielens. *American Economic Review*, 100 (4), 1358-98.
- Coopersmith, S. 1967. *The Antecedents of Self-Esteem*. San Francisco: Freeman.
- Ertac, S. 2005. Social Comparisons and Optimal Information Revelation: Theory and experiments. Working Paper.

- Ewenstein B., B. Hancock, A. Komm. 2016. Ahead of the curve: The future of performance management. *McKinsey Quarterly*, 2, 64-73.
- Eyring, H., V. G. Narayanan. 2018. Performance effects of setting a high reference point for peer-performance comparison. *Journal of Accounting Research*, 56 (2), 581-615.
- Floyd, E., J. A. List. 2016. Using field experiments in accounting and finance. *Journal of Accounting Research*, 54 (2), 437-475.
- Freixas, X., R. Guesnerie, J. Tirole. 1985. Planning under incomplete information and the ratchet effect. *Review of Economic Studies*, 52 (2), 173-191.
- Gerber, J. P., L. Wheeler, J. Suls, 2018. A social comparison theory meta-analysis 60+ years on. *Psychological Bulletin*, 144 (2), 177-197.
- Gill, D., Z. Kissová, J. Lee, V. Prowse. 2019. First-place loving and last-place loathing: How rank in the distribution of performance affects effort provision. *Management Science*, 65 (2), 494-507.
- Hannan, R. L., R. Krishnan, A. H. Newman. 2008. The effects of disseminating relative performance feedback in tournament and individual performance compensation plans. *Accounting Review*, 83 (4), 893-913.
- Hannan, R. L., G. P. McPhee, A. H. Newman, I. D. Taftkov .2013. The effect of relative performance information on performance and effort allocation in a multi-task environment. *Accounting Review*, 88 (2), 553-575.
- Ilies, R., T. A. Judge. 2005. Goal regulation across time: the effects of feedback and affect. *Journal of Applied Psychology*, 90 (3), 453-467.
- Jung, M., D. Cho, E. Shin. 2021. Repairing a cracked mirror: the heterogeneous effect of personalized digital nudges driven by misperception. *Production and Operations Management*, 30 (8), 2586-2607.
- Kahneman, D., A. Tversky. 1979. Prospect theory: an analysis of decision under risk. *Econometrica*, 47 (2), 263–291.
- Kandel, E., E. P. Lazear. 1992. Peer pressure and partnerships. *Journal of Political Economy*, 100 (4), 801-817.
- Kluger, A. N., A. DeNisi. 1996. The effects of feedback interventions on performance: A historical review, a meta-analysis, and a preliminary feedback intervention theory. *Psychological Bulletin*, 119 (2), 254.
- Kuhnen, C. M., A. Tymula. 2012. Feedback, self-esteem, and performance in organizations. *Management Science*, 58 (1), 94-113.
- Lehmann, E. 1955. Ordered families of distributions. *Annals of Mathematical Statistics*, 26, 399-419.
- Leone, A. J., S. Rock. 2002. Empirical tests of budget ratcheting and its effect on managers' discretionary accrual choices. *Journal of Accounting and Economics*, 33 (1), 43-67.
- List, J. A., S. Sadoff, M. Wagner. 2011. So you want to run an experiment, now what. Some simple rules of thumb for optimal experimental design. *Experimental Economics*, 14 (4), 439-457.
- Lourenço, S. M. 2016. Monetary incentives, feedback, and recognition—complements or substitutes? Evidence from a field experiment in a retail services company. *Accounting Review*, 91(1), 279-297.
- Lourenço, S. M., J. O. Greenberg, M. Littlefield, D. W. Bates, V. G. Narayanan. 2018. The performance effect of feedback in a context of negative incentives: evidence from a field experiment. *Management Accounting Research*, 40, 1-14.
- Powell, D. Quantile regression with nonadditive fixed effects. *Empirical Economics*. Forthcoming.
- Sprinkle, G. B. 2000. The effect of incentive contracts on learning and performance. *Accounting Review*, 75 (3), 299-326.

- Sprinkle, G. B. 2003. Perspectives on experimental research in managerial accounting. *Accounting, Organizations and Society*, 28 (2-3), 287-318.
- Staats, B. R., H. Dai, D. Hofmann, K. L. Milkman. 2016. Motivating process compliance through individual electronic monitoring: an empirical examination of hand hygiene in healthcare. *Management Science*, 63 (5):1563-1585.
- Sun, J., D. J. Zhang, H. Hu, J. A. Van Mieghem. 2021. Predicting human discretion to adjust algorithmic prescription: a large-scale field experiment in warehouse operations. *Management Science*, 68 (2), 846-865.
- Tafkov, I. D. 2013. Private and public relative performance information under different compensation contracts. *Accounting Review*, 88 (1), 327-350.
- Tesser, A. 1988. Toward a self-evaluation maintenance model of social behavior. In L. Berkowitz, ed. *Advances in experimental social psychology*, 21, 181-227. Academic Press.
- Topkis, D. M. 1978. Minimizing a submodular function on a lattice. *Operations Research*, 26 (2), 305-321.
- Weitzman, M. 1980. The “ratchet principle” and performance incentives. *Bell Journal of Economics*, 11(1), 302-308.
- Young, A. 2019. Channeling Fisher: randomization tests and the statistical insignificance of seemingly significant experimental results. *Quarterly Journal of Economics*, 134(2), 557-598.

Figures and tables

Table 1. Summary statistics

Variable	N	Mean	Median	Std. dev.	Min.	Max.
Individual performance	1,105	95.835	93.602	23.067	22.346	193.61
Performance at next level	1,105	82.886	90	38.8356	0	130
Performance distance to next level	1,105	5.486	4.610	6.030	0	57.654
Hourly bonus	1,105	1.045	0.75	.946	0	2.5
Hourly would-be bonus	1,105	1.399	1.25	.873	.25	2.5
Bonus increase	1,105	.354	0.25	.225	0	.75
Performance quantile	1,105	65.559	69	23.281	4	100
Maximum performance	1,105	152.257	151.565	25.417	113	230.677
Distance to maximum performance	1,105	56.421	52.580	32.592	0	177.717

Note. Performance quantile and Maximum performance are defined with respect to the workers’ operation groups.

Table 2: Treatment Effects

	Model 1	Model 2	Model 3
	Log (Performance)	Log (Performance)	Log (Performance)
Panel A			
IF	.01770 (.02090)		
RF	.03370 (.02200)		
FF	.02180 (.02040)		
Distance to next level		-0.00655 ** (.00302)	
Bonus increase		0.0888 ** (.03450)	
Performance quantile		0.00047 * (.00024)	
Distance to max performance		-0.00028 (.00030)	
Pre-experiment week	.06880 ** (.03350)	0.0572 * (.02860)	0.0624 * (.03230)
Post-experiment week	.02520 (.02030)	-0.0354 * (.01830)	-0.0306 (.01980)
Individual performance max	.06940 (.06170)	0.0602 (.05380)	0.0466 (.05070)
Relative performance max	.00018 (.01430)	-0.0292 (.02650)	-0.0313 (.02750)
Distance to next level (IF)			-0.00496 (.00407)
Bonus increase (IF)			0.0932 * (.05310)
Performance quantile (RF)			0.00042 (.00032)
Distance to max performance (RF)			0.00001 (.00037)
Distance to next level (FF)			-0.00923 ** (.00436)
Bonus increase (FF)			0.0639 (.06370)
Performance quantile (FF)			0.00076 * (.00044)
Distance to max performance (FF)			-0.00014 (.00042)
Constant	4.51400 *** (.01850)	4.525 *** (.01400)	4.52 *** (.01750)
N	1105	1105	1105
P-value of joint significance F-test	.499	.088	.235
Panel B			
p-values under randomization testing			
IF	0.3923		
RF	0.1348		
FF	0.2995		
Distance to next level		0.03571	
Bonus increase		0.01094	
Performance quantile		0.05176	
Distance to max performance		0.36379	
Distance to next level (IF)			0.25249
Bonus increase (IF)			0.06288
Performance quantile (RF)			0.19861
Distance to max performance (RF)			0.98217

Note. Levels of significance: * 10-percent, ** 5-percent, *** 1-percent. Linear regressions with worker fixed effects and clustered standard errors by worker. Panel A of the table shows estimated coefficients and (in parentheses) standard errors. Panel B of the table shows the computed randomized inference p-values.

Table 3. Treatment effects addressing order effects

	Model 1	Model 2	Model 3	Model 4
	First two weeks	First two weeks		
	Log (Performance)	Log (Performance)	Log (Performance)	Log (Performance)
Distance to next level	.00269 (.00346)		-.00537 (.00348)	
Bonus increase	.18900 *** (.05770)		.10300 ** (.05110)	
Performance quantile	.00114 ** (.00054)		.00113 ** (.00050)	
Distance to max performance	.00104 ** (.00045)		.00042 (.00033)	
Distance to next level (IF)		.00276 (.00361)		-.00495 (.00443)
Bonus increase (IF)		.21400 ** (.09750)		.09090 (.08050)
Performance quantile (RF)		.00108 (.00075)		.00084 (.00061)
Distance to max performance (RF)		.00151 ** (.00073)		.00034 (.00050)
Distance to next level (FF)		.00307 (.00957)		-.00656 (.00430)
Bonus increase (FF)		.17600 (.13200)		.04060 (.06630)
Performance quantile (FF)		.00165 * (.00084)		.00167 *** (.00059)
Distance to max performance (FF)		.00044 (.00109)		.00074 (.00055)
Individual performance max	.16300 ** (.06570)	.16300 * (.08320)	.08650 (.06370)	.05940 (.06110)
Relative performance max	-.06660 (.07680)	-.07240 (.09200)	-.02550 (.02710)	-.03130 (.03070)
Pre-experiment week	.15700 ** (.05940)	.16500 ** (.07490)	.08890 (.05500)	.08270 (.05900)
Post-experiment week	.06730 (.04550)	.07500 (.06070)	-.00395 (.04370)	-.01020 (.04690)
Order IF			.00994 (.01060)	.00893 (.01100)
Order RF			-.01760 (.01100)	-.01100 (.01360)
Order FF			-.02800 ** (.01200)	-.03630 *** (.01140)
Order Control			.01010 (.01420)	.00820 (.01490)
Constant	4.42200 *** (.04820)	4.41400 *** (.06400)	4.49300 *** (.04470)	4.49900 *** (.04800)
N	797	797	1105	1105
P-value of joint significance F-test	0.0069	0.0062	0.025	0.013

Note. Levels of significance: * 10-percent, ** 5-percent, *** 1-percent. Linear regressions with worker fixed effects and clustered standard errors by worker.

Table 4. Quantile Regressions of Treatment Effects

	Model 1	Model 2	Model 3
	Q25	Q50	Q75
Distance to next level	-.00751 *** (.00085)	-.01040 *** (.00090)	-.00425 *** (.00089)
Bonus increase	.11800 *** (.02240)	.10200 *** (.01600)	.02810 *** (.00760)
Performance quantile	.00127 *** (.00015)	.00048 ** (.00021)	.00045 (.00061)
Distance to max performance	-.00050 ** (.00020)	-.00070 *** (.00021)	-.00081 (.00108)
Individual performance max	.18200 ** (.07540)	.12600 ** (.06280)	.29000 *** (.01850)
Relative performance max	.02980 (.06790)	.00681 (.12600)	.08160 * (.04890)
Pre-experiment week	.08980 (.07980)	.00809 (.01990)	.05440 *** (.00861)
Post-experiment week	-.06330 (.08380)	-.21000 *** (.05750)	.04170 * (.02490)
N	1,105	1,105	1,105
p-value of joint significance F-test	.000	.000	.000

Note. Levels of significance: * 10-percent, ** 5-percent, *** 1-percent. Estimations are based on Powell's (forthcoming) quantile regression for panel data (QRPD) method with adaptive Markov Chain-Monte Carlo optimization. The table shows estimated coefficients and (in parentheses) standard errors.

Figure 1: Sample distribution of individual performance

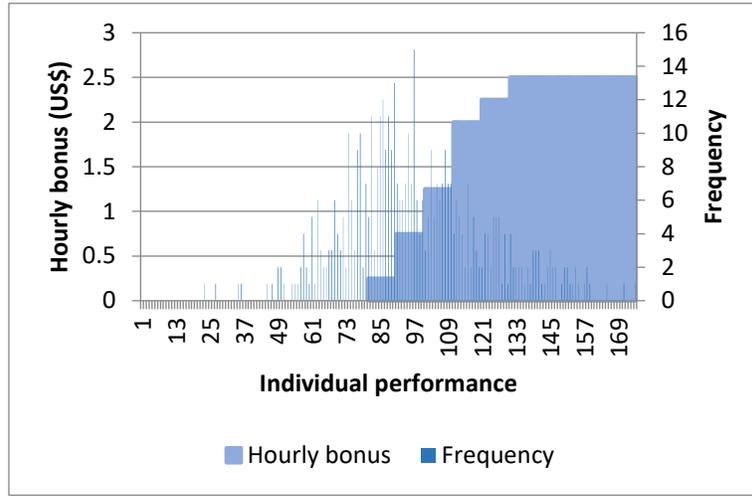
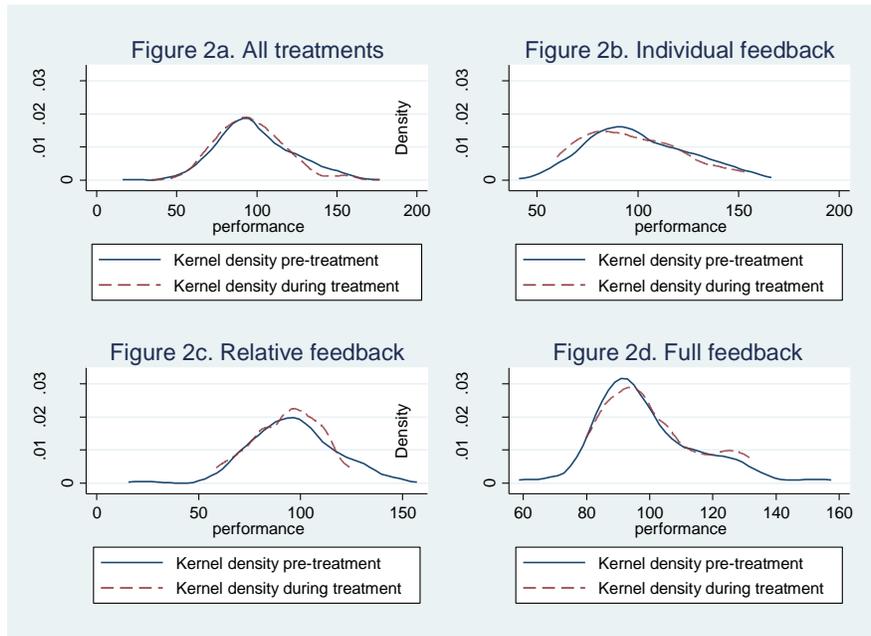


Figure 2. Performance distribution pre-treatment and during treatment



**Electronic Companion- Performance Feedback and Productivity:
Evidence from a Field Experiment, by Amrou Awaysheh, Rocio Bonet
and Jaime Ortega**

This e-companion is intended to present supplementary information to the paper. We organize the information according to the each of the paper sections for which we provide supplementary information. The number of each e-companion section is preceded by EC.

EC1. Supplementary information for Section 2. Related literature.

Table 1 provides a summary of the relevant empirical literature organized by the main aspects in which we center our background literature discussion.

Table 1. Related Literature

Study	Estimates effect of feedback content¹	Feedback on marginal effects²	Distributional effects³	Quantile estimates⁴	Monetary incentives⁵	Empirical method	Receivers of feedback	Feedback type	Performance variable
Ilies and Judge (2005)	No	No	No	No	None	Laboratory experiment	University students	Individual and relative performance	Self-reported performance goal
Hannan et al. (2008)	No	No	No	No	Tournament and Individual incentive pay	Laboratory experiment	University students	Relative performance	Profit points
Casas-Arce and Martinez-Jerez (2009)	Yes	No	No	No	Tournament	Natural experiment	Retail stores	Relative performance	Sales
Azmat and Iriberry (2010)	No	Yes	Yes	Yes	None	Natural experiment	High school students	Relative performance	Grades
Chen et al. (2010)	Yes	No	Yes	No	None	Randomized field experiment	Users of a movie recommender site	Relative performance	Productivity
Blanes i Vidal and Nossol (2011)	No	No	Yes	No	Individual incentive pay	Natural experiment	Warehouse employees	Relative performance	Productivity and quality
Kuhnen and	Yes	No	Yes	No	None	Laboratory experiment	University students	Relative performance	Problems solved

Tymula (2012)									
Tran and Zeckhauser (2012)	No	No	Yes	No	None	Randomized field experiment	University students	Relative performance	Exam grade
Hannan et al. (2013)	No	No	No	No	Effort-based pay	Laboratory experiment	University students	Relative performance	Problems solved
Kolstad (2013)	Yes	No	Yes	No	Patient demand	Natural experiment	Surgeons	Individual and relative performance	Patient mortality
Tafkov (2013)	No	No	No	No	Individual incentive pay	Laboratory experiment	University students	Relative performance	Time spent, problems solved
Newman and Tafkov (2014)	Yes	No	Yes	No	Tournament	Laboratory experiment	University students	Relative performance	Profit points
Casas-Arce et al. (2017)	Yes	No	Yes	No	Individual incentive pay	Randomized field experiment	Home repair professionals	Individual performance	Customer satisfaction and professionals' input
Eyring and Narayanan (2018)	No	No	Yes	No	None	Randomized field experiment	Online course students	Relative performance	Student activity level and grade
Lourenço et al. (2018)	No	No	Yes	No	Threat of contract termination	Randomized field experiment	Physicians	Individual performance	Electronic prescription rate
Azmat et al. (2019)	No	No	Yes	No	None	Randomized field experiment	University students	Individual and relative performance	Grades, self-reported effort and satisfaction
Gill et al. (2019)	Yes	No	No	No	None	Laboratory experiment	University students	Relative performance	Number of words

									spotted, number of correct calculations
Jung et al. (2021)	Yes	No	Yes	No	Cost saving	Randomized field experiment	Electricity customers	Individual performance (energy consumption)	Energy consumption

Notes. (1) This column indicates whether the study estimates how the effect of feedback varies according to its content, i.e. according to the value of individual or relative performance that is transmitted to the feedback receiver. (2) This column indicates whether the feedback includes information related to the effect that an increase in the receiver's performance would have. (3) This column indicates whether the study provides estimates of the effect of feedback on the distribution of performance. These may be explicit estimates of distributional effects or estimates that are suggestive of distributional effects. (4) This column indicates whether distributional effects are estimated using quantile regressions. (5) This column indicates whether there is a link between feedback receivers' performance and pay.

EC2. Supplementary information for section 5. Experimental design

A. *Experimental groups and order of the treatments*

Table 2A shows a detailed description of the 2x2 design we employed, assigning these three treatments of feedback and the control across the four different sections of the. Table 2B shows the order with which after two weeks, each group was rotated.

Table 2A: Experimental groups

		Relative feedback	
		NO	YES
Individual feedback	NO	Control	Treatment group 2
	YES	Treatment group 1	Treatment group 3

Table 2B: Order of treatments

Section	Treatment group			
	Weeks 1-2	Weeks 3-4	Weeks 5-6	Weeks 7-8
1	IF	RF	FF	Control
2	RF	FF	Control	IF
3	FF	Control	IF	RF
4	Control	IF	RF	FF

Note. IF = Individual Feedback; RF = Relative Feedback; FF = Full Feedback

B. *Templates used in the experiment*

The standard templates that were used to create the messages given to each individual employee were the following:

Control Group Message:

«First_Name» «Last_Name»

Remember Safety is our priority. As you carry out your activities in the warehouse please keep in mind your safety and the safety of those around you. If there is anything related to safety that you have a concern with, please contact your supervisor immediately. Have a safe day!

Treatment Group 1 Message:

«First_Name» «Last_Name»

Your average performance last week was «performance» %, based on that you earned a bonus of \$«bonus» per hour. If you achieved an extra «distance_from_next_level»% in your performance, you would have earned a bonus of \$«wouldbebonus» per hour.

Remember Safety is our priority. As you carry out your activities in the warehouse please keep in mind your safety and the safety of those around you. If there is anything related to safety that you have a concern with, please contact your supervisor immediately. Have a safe day!

Treatment Group 2 Message:

«First_Name» «Last_Name»

Your average performance last week was «performance»%. «worsepercnnumber»% employees of the employees in your area had a lower performance than you, and the best performer did «operationmaxperf»%.

Remember Safety is our priority. As you carry out your activities in the warehouse please keep in mind your safety and the safety of those around you. If there is anything related to safety that you have a concern with, please contact your supervisor immediately. Have a safe day!

Treatment Group 3 Message:

«First_Name» «Last_Name»

Your average performance last week was «performance»%, based on that you earned a bonus of \$«bonus» per hour. If you had achieved an extra «distance_from_next_level»% in your performance level, you would have earned a bonus of \$«wouldbebonus» per hour.

«worsepercnnumber» % employees in your area had a lower performance than you, and the best performer did «operationmaxperf» %.

Remember Safety is our priority. As you carry out your activities in the warehouse please keep in mind your safety and the safety of those around you. If there is anything related to safety that you have a concern with, please contact your supervisor immediately. Have a safe day!

C. Differences across sections:

Before running the experiment, we explored differences among the four sections using t tests performing dyadic comparisons across sections and found no significant difference in terms of average performance. These differences are shown in Table 3A. Furthermore, a Krustal–Wallis test of equality of populations failed to reject the hypothesis of equality of the performance distribution across the four sections ($p=0.844$). When comparing information about other relevant characteristics: distance to next level, bonus increase, performance quartile, and distance to max

performance (information we later provided during our experiment), we found some significant differences across the sections. Specifically, we find statistical significance for the amount that the bonus would increase if they were to perform to the next level in the pay-for-performance pay scheme, for the distance to the maximum performance of the section, and for their performance quantile. Given these differences, the rotation of the treatments may help in assuaging concerns that any effects that we find are due to differences across the sections.

Table 3A. Average section characteristics pre-experiment

	Section 1	Section 2	Section 3	Section 4
Pre-experiment performance	100.363	96.360	99.210	100.632
Permanent employees	21	16	8	9
Performance distance to next level	5.360	5.702	4.492	6.587
Bonus increase	0.359	0.416*	0.462***	0.314***
Distance to maximum performance	57.632***	41.125***	45.801	56.365*
Performance quantile	65.404	69.614*	67.962	58.862**

Note. *10-percent, **5-percent, ***1-percent levels of significant difference regarding differences between the average of the variable at the section against all the other sections.

EC3. Supplementary information for section 6. Data and Variables

Table 3B shows detailed information regarding the composition of the sample.

Table 3B. Sample composition

Group	Percent	N	Weeks
Pre-experiment	29.68	328	6
Individual feedback treatment	9.59	106	8
Relative feedback treatment	9.68	107	8
Joint treatment	9.23	102	8
Control group	9.05	100	8
Post-experiment	32.76	362	8
Total	100.00	1,105	22

Table 4 shows how the main empirical variables relate to the theoretical variables of the model and Table 5 shows the name and definition of the variables used in the main empirical analyses.

Table 4. Theoretical variables and empirical proxies

Theoretical variables	Empirical proxies	Explanation
Productivity (x_i)	Individual performance	Productivity is defined as the time that the worker should have spent to complete the tasks (according to the standard) divided by the time that the worker actually spent.

Bonus pay (βx_i)	Hourly bonus	The hourly bonus is defined as the amount of dollars per hour according to the company's bonus scheme.
Feedback on bonus pay (z_i)	Bonus increase Performance distance to next level	Since the bonus scheme is not linear the marginal return of effort is not constant and cannot be summarized with one single variable. To give workers information about the marginal return of effort we need to at least tell them how close they are to the next level and also what bonus increase they would earn if they reached the next level.
Distance to maximum productivity ($x_i - \max(x_i)$)	Distance to maximum performance	This distance is the actual difference between the worker's performance and that of the top performer.
Feedback on distance to maximum productivity (z_i)	Distance to maximum performance Performance quantile	These variables constitute two alternative ways that we use to give feedback about the employee's distance to the top performer. Distance to maximum performance is the actual distance in productivity units and the performance quantile is an alternative measure in percentage terms.

Table 5. Main variables names and definitions.

Variable Name	Definition
Log Individual Performance	Total time it should have taken an employee to complete all the tasks assigned divided by the actual time spent to complete those tasks. We take the log of performance because performance is skewed to the right.
IF	Dichotomous variable =1 if the individual performance feedback was given, =0 otherwise
RF	Dichotomous variable =1 if the relative performance feedback was given, =0 otherwise
FF	Dichotomous variable =1 if the individual and relative performance feedback were given at the same time, =0 otherwise
Distance to next level	Performance distance to the immediately higher bonus threshold in the previous week
Bonus increase	Difference between the bonus corresponding to the next level of performance and the actual bonus earned in the previous week
Performance quantile	Percentage of individuals in the section that performed worse in the previous week
Distance to max performance	Performance distance to the top performer of the section in the previous week
Distance to next level (IF, FF)	Variable =Distance to next level if the employee received the IF or FF treatment, =0 otherwise
Bonus increase (IF, FF)	Variable =Bonus increase if the employee received the IF or FF treatment, =0 otherwise
Performance quantile (RF, FF)	Variable =Performance quantile if the employee received the RF or FF treatment, =0 otherwise
Distance to max performance (RF, FF)	Variable =Distance to max performance if the employee received the RF or FF treatment, =0 otherwise
Distance to next level (IF)	Variable =Distance to next level if the employee received the IF treatment, =0 otherwise
Bonus increase (IF)	Variable =Bonus increase if the employee received the IF treatment, =0 otherwise
Performance quantile (RF)	Variable =Performance quantile if the employee received the RF treatment, =0 otherwise
Distance to max performance (RF)	Variable =Distance to max performance if the employee received the RF treatment, =0 otherwise
Distance to next level (FF)	Variable =Distance to next level if the employee received the FF treatment, =0 otherwise
Bonus increase (FF)	Variable =Bonus increase if the employee received the FF treatment, =0 otherwise
Performance quantile (FF)	Variable =Performance quantile if the employee received the FF treatment, =0 otherwise
Distance to max performance (FF)	Variable =Distance to max performance if the employee received the FF treatment, =0 otherwise
Individual performance max	Dichotomous variable =1 if the individual reached the maximum level of bonus, =0 otherwise
Relative performance max	Dichotomous variable =1 if the individual was the top performer in her section, =0 otherwise
Pre-experiment week	Dichotomous variable =1 if the individual-week observation corresponds to the weeks before the experiment, =0 otherwise
Post-experiment week	Dichotomous variable =1 if the individual-week observation corresponds to the weeks after the experiment, =0 otherwise

EC4. Supplementary information to Section 7. Results

A. *Feedback effects on performance in the workplace: descriptive results*

To explore the average of the treatment effects, we also compare the average performance of each section for each of the treatments with the average performance of the corresponding section before the experiment (see Table 6). That is, for each section, we compare the average performance associated with the two weeks of each intervention (IF, FF, RF, Control) with the average performance before the experiment started in that same section. With the exception of the control intervention in section 3 (weeks 3 and 4), for which we find average performance during these two weeks (88.57) to be significantly lower than average performance before the experiment started (99.21), these dyadic comparisons do not reveal any significant differences.

In addition, we also compare, within a given intervention (IF, RF, FF, Control) the average performance across the sections in a dyadic way (i.e., we perform all dyadic comparisons possible, e.g., average performance of IF in section 1 versus average performance of IF in section 2, average performance of IF in section 1 versus average performance of IF in section 3, and average performance of IF in section 1 versus average performance of IF in section 4, etc.). These data do not show a clear trend of performance average differences depending on when the information was given (earlier versus later weeks), which suggests that the order effects may not be that strong in here.

Table 6: Average Performance by Treatment and by Weeks

TREATMENT	Weeks 1-2	Weeks 3-4	Weeks 5-6	Weeks 7-8
IF	97.93 (S1)	97.64 (S4)	94.94 (S3)	90.81 (S2)
RF	93.23 (S2)	103.25 (S1)	89.87 (S4)	93.74 (S3)
FF	100.46 (S3)	92.49 (S2)	100.49 (S1)	91.86 (S4)
Control	91.39 (S4)	88.57 (S3)**	92.79 (S2)	100.07 (S1)

Note. **Significant difference at 5 percent level between pre-experiment performance and performance during the treatment weeks.

B. *Additional analyses*

We perform a series of additional analyses to provide suggestive evidence of our main mechanism as well as to explore the robustness of our results to alternative specifications. In here we explain in detail how we do that.

First, we provide evidence of the learning mechanism that we hypothesize in our model by exploring the extent to which absolute and relative performance are stable over time and whether this persistence changed with the experiment. If an employee is not truly receiving any new information with the experiment, we do not expect the persistence of performance to change with

the feedback we give them. To do this, we regress workers' performance (individual and rank position) in a focal week on their previous week performance (individual and rank position) and on the interaction between the experiment period and the previous week's performance (individual and rank position). The results are shown in Table 7. We find that the coefficient associated with the interaction term (Previous individual performance \times Experiment period) is negative and significant, suggesting that the relationship between prior performance and performance became weaker during the experiment period, which is consistent with learning effects. However, the interaction term (Previous relative performance \times Experiment Period) is not significantly different from zero for relative performance. The findings regarding relative performance are similar to those in Blanes i Vidal and Nossol (2011), who found that even though relative feedback was increasing performance overall, the relative ranking of individuals remained unchanged.

Table 7. Effects of previous week performance on current performance, individual performance (Model 1) and relative performance (Model 2)

	Model 1	Model 2
	Individual	Relative performance
Previous individual performance	.816 *** (.024)	
Experiment period	12.760 ** (5.917)	5.027 (3.422)
Previous individual performance * Experiment period	-.127 ** (.061)	
Previous relative performance		.770 *** (.038)
Previous relative performance*Experiment period		-.071 (.048)
Constant	16.880 *** (2.285)	14.720 *** (2.468)
N	1,039	1,039

Note. Levels of significance: * 10-percent, ** 5-percent, *** 1-percent. The table shows estimated coefficients and (in parentheses) standard errors

Next, we consider whether a Hawthorne effect could be an alternative explanation for our results. To do that, we re-estimate our models by removing the data from the two weeks immediately before the experiment started, when participants learned that the study was going to take place and could have already reacted to the experimenter introduction. When excluding these data, we continue to find equivalent results (analyses available upon request).

In addition, we explore whether our results are robust to alternative specifications (analyses available upon request). First, we use an alternative operationalization of the information variables. Specifically, we take into account the fact that bonus increases are nonmonotonic in performance and explore whether workers react to performance feedback differently, specifically to individual performance feedback, when they are in the incentive zone (i.e., the zone in which

increasing performance would lead to a change in bonus) versus in the flat zones (either in the floor zone because performance is less than 80 percent or in the ceiling zone because performance is greater than 130). We estimate our model replacing the information variables with three dichotomous variables that capture the zone where workers are located (incentive, floor, ceiling), which workers could infer when they received individual performance information. In these regressions we find that only the variable incentive zone has a positive and significant coefficient on performance, which reinforces our conclusion that learning that you are far from getting a bonus or have reached the ceiling of the bonus scheme is associated with no performance increase. Second, we also explore the robustness of our results to an alternative specification. In our main analyses we have included post-experiment performance to control for potential across the unit overall performance time trends. If there was a general trend towards lower performance, this should be captured by the pre- and post- experiment dummies. It seems that performance of the control group was higher on average in the period before receiving information and was lower after the experiment concluded. As an alternative specification of the control variables structure, we remove the variable post-experiment performance. Importantly, we continue to find significance for all of the variables in Model 1 with the exception of Performance quantile (that used to be significant at 10%). Moreover, although the coefficients continue to be of the same sign and similar magnitude, we lose significance in the main variables of Model 2. The lack of statistical significance may be due to the drop in the sample size.

We also explore whether our results are robust to an alternative clustering of the standard errors. Given that workers were located in different sections, there is a concern that standard errors might vary across both workers and sections. To account for that, we repeat our analyses clustering the standard errors by worker and section, and we find similar results as when we cluster only by worker.

Finally, we explore whether our results are robust when considering only the workers who stayed throughout the whole experiment period. Our main analyses were estimated using the maximum number of observations for the whole period, i.e., without considering potential turnover of permanent workers during the period of analysis. Although the turnover of permanent workers was relatively low in our setting (average of 1.02 percent weekly turnover rate across all four sections), we check whether our results are robust when excluding from the sample permanent workers who turned over or who joined the sample during the experiment. We find our results to be robust: although we lose statistical significance for two of the variables in Model 1, namely Distance to next level (that used to be significant at 5-percent level) and Performance quantile (that used to be significant at 10-percent level), the signs and magnitude of coefficients of the rest of the variables are comparable to those when not limiting the sample.