4-16-2015

The Contingent Effect of Management Practices

Steven Blader

Claudine Madras Gartenberg

Andrea Prat

Recommended Citation

Blader, S., Gartenberg, C. M., & Prat, A. (2015). The Contingent Effect of Management Practices. *Columbia Business School Research Paper No. 15-48,* Retrieved from https://repository.upenn.edu/mgmt_papers/273

The Contingent Effect of Management Practices

Abstract
This paper investigates how the success of a management practice depends on the nature of the long-term relationship between the firm and its employees. A large US transportation company is in the process of fitting its trucks with an electronic on-board recorder (EOBR), which provide drivers with information on their driving performance. In this setting, a natural question is whether the optimal managerial practice consists of: (1) Letting each driver know his or her individual performance only; or (2) Also providing drivers with information about their ranking with respect to other drivers. The company is also in the first phase of a multi-year initiative to remake its internal operations. This first phase corresponds to an overhaul of the relational contract with its employees, focusing exclusively on changing values toward a greater emphasis on teamwork and empowerment. The main result of our randomized experiment is that (2) leads to better performance than (1) in a particular site if and only if the site has not yet received the values intervention, and worse performance if it has. The result is consistent with the presence of a conflict between competition-based managerial practices and a cooperation-based relational contract. More broadly, it highlights the role of intangible relational factors in determining the optimal set of managerial practices.

Keywords
relational contracts, management practices, transportation, performance ranking

Disciplines
Management Sciences and Quantitative Methods

The Contingent Effect of Management Practices

Steven Blader — New York University
Claudine Gartenberg — New York University
Andrea Prat — Columbia University

Initial draft: October 30, 2014

This draft: September 2, 2016*

Abstract

This paper investigates how the success of a management practice depends on the nature of the long-term relationship between the firm and its employees. A large US transportation company is in the process of fitting its trucks with an electronic on-board recorder (EOBR), which provide drivers with information on their driving performance. In this setting, a natural question is whether the optimal managerial practice consists of: (1) Letting each driver know his or her individual performance only; or (2) Also providing drivers with information about their ranking with respect to other drivers. The company is also in the first phase of a multi-year initiative to remake its internal operations. This first phase corresponds to an overhaul of the relational contract with its employees, focusing exclusively on changing values toward a greater emphasis on teamwork and empowerment. The main result of our randomized experiment is that (2) leads to better performance than (1) in a particular site if and only if the site has not yet received the values intervention, and worse performance if it has. The result is consistent with the presence of a conflict between competition-based managerial practices and a cooperation-based relational contract. More broadly, it highlights the role of intangible relational factors in determining the the optimal set of managerial practices.

1 Introduction

Economists have increasingly focused on management practices as an important explanation for the large observed variation in productivity among firms (Bloom and Van Reenen 2007; Syverson 2011). This explanation logically raises the following question: why do large differences in practice adoption persist across firms, even

*We are particularly indebted to Rebecca Henderson who has provided extensive support and advice during this ongoing research program and of course to our partner company for their collaboration. We would also like to thank Nava Ashraf, Oriana Bandiera, Iwan Barankay, Roland Benabou, Nick Bloom, Wouter Dessein, Ray Fisman, Bob Gibbons, Jonas Hjort, Amit Khandelwal, Botond Kőszegi, Stephan Meier, Raffaella Sadun, Ned Smith, John Van Reenen, Eric Van den Steen, Christopher Woodruff and participants at the NBER Organizational Economics Working Group, Duke Strategy Conference, SRF Annual Meeting, AEA Annual Meeting, and seminars at NYU, Columbia, London Business School, and Middlebury.

within narrowly defined industries? One proposed answer is that poor institutions make adoption costly and protect inefficient firms from competition (Bloom and van Reenen 2010). This answer applies primarily to developing countries and leaves unaddressed the large observed dispersion in developed settings and the substantial within-country variation of practice adoption, where firms presumably operated under similar institutional contexts.

A second explanation, explored by Gibbons and Henderson (2013) and Helper and Henderson (2014), is based on relational contracts. The ability of a firm to introduce a new management practice - and whether that practice is optimal for that firm - may depend on the relational contract that is in place. A relational contract is a non-legally binding understanding between a firm and its employees that typically describes how employees should behave and how the firm will reward the expected behavior (MacLeod and Malcomson 1989; Baker et. al. 2002; Levin 2003). The contract, which is enforced through repeated interaction, may display path-dependence (Chassang 2010; Halac 2012). As a result, similar firms may be governed by very different relational contracts, which can in turn drive differences in the effectiveness of observable management practices.

This proposed influence of relational contracts on managerial practices is difficult to measure for two reasons. First, by definition, relational contracts are difficult to observe. They are implicit informal agreements rather than explicit written contracts exactly because they contain prescriptions that cannot be expressed in a legally binding way. Second, even if they were observable, given that every company has its own relational contract, it would be difficult to disentangle the effect of the relational contract from other firm-specific unobservable factors.

The goal of this paper is to make some progress on this issue by studying a company arguably transitioning from one relational contract to another. This complex transition is implemented on a site-by-site basis across the company and is a lengthy process. The company was mid-way through the transition at the time of our experiment and therefore some locations within the firm had experienced the change while others had not. We study how workers in these different locations react to the introduction of the same management practice.

Specifically, we run a field experiment within a transportation company with a large number of sites that all perform a similar function throughout the United States. Our company recently introduced electronic on-board recorder (EOBR) technology that measures the performance of drivers against a route-specific benchmark. The introduction of this technology raises a question about the optimal managerial practice for sharing performance information. In particular, should drivers be made aware only of their individual performance or should performance statistics of all drivers at a given site be posted and made publicly viewable? Both options provide performance feedback to drivers. The latter practice is also likely to spur comparisons and potentially competition between drivers; however, it is unclear how this will influence driver performance. We worked with the company's management to shed light on this issue by running a randomized controlled field study.

As mentioned above, the company is also in the midst of a major transition that involves reshaping its relational contract with employees. Specifically, they are engaged in a multi-year program to radically overhaul their business practices and culture. The company's ultimate goal is to base their operations on a "Lean Management" management philosophy, inspired by the Toyota Production System, centered on teamwork and worker empowerment (Womack et. al. 1990; Holweg 2007). Given that the firm's initial relational contract was characterized by the prevailing individualistic culture at our company, a successful implementation of Lean Management requires profound changes across all levels of the organization. Accordingly, the company committed substantial resources and set a ten-year time horizon for the implementation.

This implementation is composed of five phases and, at the time of the experiment, the company was midway through the first phase. Crucially, the first phase does not impose any change in the work processes of the drivers, their incentives or the formal management practices under which the sites operated. Rather, the first phase is primarily focused on reorienting the workplace culture towards teamwork, collective effort and the empowerment of front-line workers. This represented a significant change for employees at our company, since the previous management philosophy was based on individual behavior, limited delegation, and top-down supervision. It involved supervisors and managers taking steps to change the culture of their sites to be more open to employees taking initiative, raising problems freely, and addressing those problems as a group. For the purpose of this paper, we refer to this first phase as "Phase 1," even though Lean Management, in its fullest sense, involves many other changes to both formal and informal operating practices (which had not yet been initiated at our company). At the time of the experiment some sites had commenced Phase 1 while the others had not.

One of the most salient features of Phase 1 is the Toyota-inspired emphasis on teamwork. Indeed, in a series of interviews we conducted, employees consistently emphasized a marked increase in "cohesion", "camaraderie", and "respect" following the introduction of Phase 1. Appealing to Benabou and Tirole's (2003) model of intrinsic motivation, we can interpret these changes as a shift in the reference structure that underlies the worker's intrinsic motivation. Specifically, since Phase 1 emphasizes a collective orientation toward work, rather than maximizing individual job satisfaction, now each worker also considers his or her team member's satisfaction. Intuitively, this leads to the prediction that the initiation of Phase 1 will reduce the effectiveness of any management practice that relies on competition between workers.

We develop this intuition formally in the theory section, through an extension of Benabou and Tirole (2003). The model analyzes two changes: a shift from private performance feedback to public performance feedback and a modification of terms of the relational contract between the firm and its workforce, in which the firm announces an emphasis on collective performance. The main result of the analysis is that introducing public feedback leads to higher performance only if the relational contract emphasizes individual, rather than collective, performance.

The primary empirical finding in our study is that the effect of posting drivers' performance strongly depends on whether the sites had received the Phase 1 intervention. Drivers assigned to untouched sites responded on average positively to public performance postings, improving their fuel efficiency by 4.5% and reducing their idling time and wasted fuel by 1.1% and 1.8%, respectively, relative to the control group. In contrast, drivers at Phase 1 sites responded negatively to the individual performance postings. We record a substantial drop in performance for these drivers, in the form of a 10.7% reduction in fuel efficiency and an increase of 2.5% in idling time and 4.4% in wasted fuel, relative to the control group.

This finding must be interpreted in light of the fact that the Phase 1 intervention did not change any existing incentives (or compensation of any kind) or formal processes. As such, a researcher who had complete site-by-site 'hard' information about the current managerial practices – but no knowledge of the fact that certain sites had been exposed to a cultural intervention – would have missed a key source of site-level adoption success.

Since Phase 1 primarily consisted of introducing employees to collaboration, teamwork and empowerment—and in communicating the role of these elements within the company on an ongoing basis – our experimental results highlight the importance of accounting for 'relational' factors by researchers seeking to understand employee behavior. In turn, these intangible relational factors (which we explore in detail later) were by definition under the control of the firm, as the Phase 1 intervention was a deliberate choice of the company. The experiment therefore indicates that the optimal set of managerial practices depends on the relational contract that the company seeks to establish with its workers.

Once we establish the main result of our experiment, we then probe in two directions: the randomness of the Phase 1 assignment and the proposed mechanism driving the experimental results.

Regarding Phase 1 assignment, while we directly randomized the performance posting treatment, we relied on the company's pre-existing rollout of Phase 1. This reliance resulted from a constraint of the experiment, which itself arose because of differences in nature of the two interventions. The company's management required the postings to be rolled out across all sites during a specific four month window at the end of 2013, while the Phase 1 intervention was scheduled to roll out across all sites over a five year period (our study was conducted in the middle of this period), with a minimum of three to six months to complete at any given site. Given this timing mismatch, we stratified our randomization of performance postings by whether a site had received the Phase 1 intervention at least three months prior to the commencement of the study. We test the validity of the randomization assumption in two ways. Following Oster (2014) and Altonji et al (2005), we first quantify the magnitude of unobservables bias and find that it far exceeds the bounds of reasonableness suggested by the authors. Second, we perform a more traditional propensity-score approach and replicate our results on the matched sample. In general, our results either remain stable or strengthen with the inclusion of controls and after the matching.

The second direction that we probe is the psychological mechanism underlying our experimental results. To do so, we analyze employee attitudes that the company collected through an annual engagement survey. As one would expect, Phase 1 sites score higher on the survey questions that assess workers' collectivistic orientation. When we replace the Phase 1 intervention dummy with a survey-based index of collectivist orientation in our primary triple-differences analysis, we replicate the pattern of results as our Phase 1 indicator. No significant pattern is observed if instead we use a different index of employee attitudes, one that focuses on individual satisfaction with compensation and benefits. These results are consistent with our explanation that the differential effects are most accurately attributed to collectivistic orientation and not to individual satisfaction.

Moreover, two additional tests shed further light on the nature of the backlash we find against public performance postings and provide additional support for our reasoning. First, we find that the triple-difference result extends to second moments in that performance posting increases performance variance in pre-Phase 1 sites and decreases it in Phase 1 sites. This result in line with our theoretical predictions: when performance postings are introduced in Phase 1 sites, top performers reduce their effort out of deference to their lower-performing teammates' satisfaction, which is harmed when performance differences become widely known. This prompts a reduction in the variance of performance within sites. This finding is also consistent with research in social psychology that finds that people in highly collectivist environments tend to converge in their thoughts and behaviors due to their concerns about the social dynamics within the group (Brown and Turner 1981; Tajfel and Turner 1986; Blader and Tyler 2009).

Second, we compare the outcomes of two different posting treatments. In both cases all drivers' scores are publicly posted, but in one case names are revealed (named postings) and in the other case they are replaced with employee IDs (IDed postings), which effectively anonymizes the results. These two conditions enable us to isolate more precisely the effects of explicit competition among employees, since they hold constant relative performance feedback and vary only in the identifiability of peers' performance. We find that only the named postings treatment matters. This finding is consistent with social psychological research showing that the competitive behavior arising from the postings should be greatly reduced when one does not know the identities of one's adversaries (Haran and Ritov 2014).

To the best of our knowledge, this is the first randomized controlled trial that shows how the successful adoption of a management practice depends on employee values that can be modified by the firm. As such, it highlights the importance for firms of making joint long-term decisions about corporate values and management practices. It makes several contributions to literatures within and outside organizational economics.

First, this link between management practices and relational contracts is conjectured, but not demonstrated, in prior discussions of practice adoption (e.g., Gibbons and Henderson, 2013 and Helper and Henderson, 2014).

It is also discussed by Ichniowski, Shaw and Prennushi (1997), who suggest that differences in practice adoption between steel finishing lines can be partly explained by differences in levels of trust between labor and management. We further this discussion by providing experimental evidence that different relational contracts can in fact shape how workers respond to the introduction of new management practices and that a practice that is beneficial under one relational contract can be detrimental under another one.

In that sense, our work also expands the discussion on complementarities between management practices (Milgrom and Roberts, 1990, 1995) in several ways. First, the formal theory and empirical tests have generally focused on "hard" choices, such as technology adoption, incentives, employee skills. For example, within the transportation industry, Hubbard (2000) and Baker and Hubbard (2003) have documented the presence of design complementarities between monitoring technology, incentive provision, and asset ownership. In our setting, importantly, there is no heterogeneity between units in incentives, asset ownership or other hard practices beyond our performance posting intervention. Beyond transportation, while there have been a number of other empirical tests of the complementarity hypothesis (for example, Ichniowski and Shaw 2003; see Brynjolfsson and Milgrom 2013 for a survey), ours is the first field experiment that explores the complementarity between management practices and intangible aspects of the worker-firm relationship. Also, we find negative interactions, which are not part of the general theory. So, while our findings are not covered in a standard complementarity setting, they suggest a way to extend Milgrom and Roberts' model both to encompass intangibles and also to allow for negative complementarities between otherwise "high performance" workplace design choices.

Our work also relates to experiments on the effect of relative feedback interventions (see for instance Bandiera et. al. 2012; Barankay 2012; Ashraf et. al. 2014; Bursztyn and Jensen 2015). In most of the existing experiments, compensation is directly based on the performance measure that forms the object of the relative feedback experiment, with Ashraf et al (2014) as one exception. Like that study, our drivers do not have an explicit incentive scheme and promotions are typically seniority based. The results of these prior studies are mixed, with some showing an improvement in performance (Blanes i Vidal and Nossol 2011; Delfgaauw et al 2012) and others showing a decline (Bandiera et. al, 2012; Ashraf et al 2014; Bursztyn and Jensen, 2015) The key contribution of our study is to propose one potential explanation for the differences in the findings; namely, that the individual response to public social comparison information may depend on the relational contract between the individuals and the organization.

This paper is also related to research that has noted two important trends in workforce practices over the past three decades. The first trend is the adoption of "innovative" human resource management practices, particularly a trend toward team-based management and group incentives (Lawler et. al. 1995, 2001 and Lawler and Mohrman 2003), perhaps reflecting increased diffusion of Japanese management practices. The second parallel trend has been the increased use of data-driven management, in which firms implement technologies that enable much closer monitoring along (some) key output factors (e.g., Lemieux et al 2009; Cowen 2013). Our paper shows that these two trends, while potentially complementary, have complex interactions that can affect the returns to firms attempting to adopt both.

The mechanism we identify is also related to intrinsic motivation in team problems. A number of empirical studies have examined the effectiveness of group incentives. Contract theory has incorporated intrinsic motivation in incentive problems though a variety of conceptual frameworks (Kandel and Lazear 1992; Koszegi 2013). Our findings most closely support the mechanism proposed in Benabou-Tirole's (2003) model of worker type signaling, which we explore in detail in the theory.

Finally, this paper highlights the need for additional caution when interpreting the result of a randomized controlled trial within an organization (see Bandiera, Barankay and Rasul 2011 for a survey). The results of a field experiment conducted in one organization only extend to other organizations with similar observables – or to the same organization at a later time – if it can be argued that those other organizations have similar relational contracts. To illustrate this point, while our results show that performance postings benefit a site if and only if the site had not initiated the Phase 1 intervention, if we had run the same experiment in 2012 (before the initiation of Phase 1), we would have found overwhelming support for performance postings. Conversely, if we had conducted our study after all sites had initiated Phase 1, we would arguably reach the opposite conclusion. Indeed, when the company's management saw our findings they decided not to publicly post driver performance because they realized that this management practice went against the relational contract they were attempting to implement.

Section 2 provides background information on the research setting. Section 3 presents a stylized model, with a number of testable predictions Section 4 describes the nature of our experiment. Section 5 reports the main results, while Section 6 contains the additional tests we perform to further examine the mechanism that underlies our main results. Section 7 concludes.

2 The Research Setting

2.1 Why the Transportation Industry?

The US transportation industry has several features that make it well suited for research on relational contracts and management practices. Intense competition and well developed information markets (in the form of trade organizations, conferences and consultants) lead firms to rapid adoption of productivity-enhancing technology. Recently, a subset of this technology has provided managers with extensive data and monitoring capabilities, enabling them to implement a broad range of previously infeasible management practices. In fact, managers are effectively required to adopt these new practices, given that the technology is only useful insofar as it is effectively integrated into the daily operations of the company. In this sense, we can view these new technologies as a shock to management practices across the industry.

One of these technologies is of particular importance to our research design. Electronic on-board recorder systems (EOBR) record and transmit detailed driving behavior to a centralized database accessible to managers. This database can be used to evaluate, discipline and reward drivers in near-real time. EOBR systems also include terminals installed in truck cabs that display driver performance information and emit audible real-time alarms when driving behavior is out of system bounds.[1]

A second feature of this setting is that the new technology and associated practices can be viewed by drivers as highly intrusive (in fact, at the time of our writing this study, a new technology was announced to install cameras that measure the height of drivers' eyelids to gauge their fatigue[2]). If implemented improperly, firms run the risk of alienating their workforce, which can result in reduced productivity, sabotage and greater union activity. From our discussions with company management, driver acceptance was a primary concern as they decided when and how to roll out new technology. In this sense, the new operating practices can be viewed as complementary to the relational contracts between managers and drivers at these companies.

Related to this point is that the industry has a long history of driver independence: companies have traditionally allowed a high degree of independence to these "last American cowboys," in exchange for long hours and monotonous work.[3] In this sense, the wave of new technology represents a challenge to this tradition, as does Lean Management's focus on teamwork and collective identity, and companies are faced with how to handle both of these innovations successfully.

2.2 The Company

The company at which we conducted this study operates in the less-than-truckload segment of the industry, transporting shipments that are smaller than full truckload freight and larger than individual parcels. At the time of our experiment, the company employed a substantial number of drivers across sites distributed throughout the US and Canada.[4] Important for our experimental design, most drivers operate local routes and there is little communication between sites. All the drivers in our company are hourly employees and none are

[1] See Baker and Hubbard (2003) for a study of the first wave of this technology, then called On Board Computing (OBC). The industry's interest in EOBRs is both the result of regulatory pressure and commercial motives (Koeth 2013). EOBRs are available for many purposes, including safety monitoring, route management, vehicle diagnostics, etc. One of their key potential benefits is to enable "fuel management and fuel use monitoring to improve controls and reduce cost."

[2] http://www.bloomberg.com/news/2014-10-07/droopy-eyelid-detector-one-solution-to-truck-crashes.html

[3] [Trucker culture has been defined by] "the sense of fierce independence, counter-cultural defiance, and unapologetic masculinity...truckers very much valued (and continue to value) not being confined within the four walls of a factory or an office" http://freakonomics.com/2009/02/27/ask-an-economist/)

[4] The actual number of sites and drivers has been removed for confidentiality purposes, although we discuss the numbers used in our study below.

contracted owner-operators (unlike the drivers studied in Baker and Hubbard (2003)).

Shipments are picked up and delivered during regular business hours via local routes of less than 300 miles covered by drivers who can serve the same customers over months or years. Intercity shipments are transported between sites via by a minority of “line-haul” drivers, typically on an overnight shift. Because of the difference in shift schedules and the small proportion of line-haul drivers, these drivers have limited interactions with the majority "city" drivers. Moreover, the switching of city drivers between sites occurs rarely; therefore, the threat of cross-contamination between sites during our study is limited. This feature enables us to establish a credible control group and two distinct treatment groups in the experiment.

The company was engaged in two major initiatives at the time of our study that we incorporated into our research design. First, beginning in August 2013 and continuing over a four month period, EOBR was rolled out for the first time to all trucks. This rollout represented the first time that company managers had information on individual driver’s efficiency and they were sensitive to how the use of this data would be accepted by the workforce. Accordingly, they were open to experimentation on certain practices as a means to decide how to integrate the technology into daily operations.

Second, beginning in 2011 and continuing during our study period, the company was engaged in a decade-long program to change their business culture and operations to conform to Japanese manufacturing practices. At the time of our study, the company had initiated Phase 1 of a five phase transition to this “lean” operations model, with a plan to complete the first phase across the remaining sites by the end of 2015.

2.3 Phase 1 of the Lean Initiative

The company divided their ten-year lean initiative into five phases. At the time of our study, only the first phase had been initiated, and only across less than half of the sites. Phase 1 was designed primarily to set the stage for the later adoption of lean-inspired processes by instilling in workers an appreciation for the principles of teamwork and empowerment. No formal processes related to the drivers were imposed as part of this initial phase. Instead, drivers at each site went through training on the ideals and culture of lean manufacturing. This meant, among other activities, having drivers, rather than managers, run meetings and work together to reorganize the community area and dock as they chose. Appendix Figure 2 shows the criteria by which sites are evaluated after completing this initial phase and reflects the emphasis on “soft” changes, such as the nature of the employee-manager relationship and the value placed on teamwork at the site. Appendix Figure A3 shows excerpts from interviews with drivers and supervisors on the impact of Phase 1 of the lean initiative at their sites. These excerpts indicate that, while workers noticed very few formal changes, they did have a strong sense that this was a radical and lasting change in workplace relations and that the degree of teamwork and the nature of the management style had both changed as a result.

The timing of this initiative had two advantages for our study. First, the first phase of this transition focused on changes to the prevailing relational contract and not on changes to formal driving practices that could otherwise affect our performance measures. The planned second through fifth phases do focus on the formal tools side of lean manufacturing, but importantly, none of these phases had been initiated at the time of our study.

Second, 35% of the sites had received the Phase 1 intervention at least three months prior to the beginning of the study, enabling a meaningful comparison between sites that had undergone the initiative and sites without any culture shift.

Third, after an initial pilot phase, the rollout of Phase 1 was scheduled around the pragmatic concern of simplifying the travel schedule of the various managers in charge of training. We consider the rollout, therefore, to be quasi-random for our purposes, in the sense that the rollout schedule is unrelated to the anticipated success of the initiative or other factors that may influence the reactions to the performance postings. We recognize the importance of this assumption and test it below.

3 A Model of Complementarities between Phase 1 and Performance Posting

This section proposes a highly stylized theoretical analysis of the effect of a change in management practices (the adoption of performance posting) and a change in the relationship between the company and its workforce which in turn changes the reference points of workers (the cultural transition introduced in Phase 1). The goal of the model is to make predictions on how employee behavior changes when the two changes are introduced either separately or jointly.

In the case of performance postings, the company makes all individuals' performance in a certain activity observable to all workers. The main idea here is that – absent other considerations – people enjoy publicly outperforming their peers and are embarrased if they publicly underperform. To capture this effect, let u_i be the direct job satisfaction of worker i. We assume that

$$u_i = y_i + b\bar{y}_{-i} + \rho\left(y_i - \bar{y}_{-i}\right) - \frac{1}{2}c_i y_i^2;$$

where: y_i is the performance of agent i, $\bar{y}_{-i}$ is the average performance of the rest of the n-person team, namely

$$\bar{y}_{-i} = \frac{\sum_{j \neq i} y_j}{n-1};$$

ρ is a parameter that captures the observability/salience of performance posting; and c is a cost parameter. Thus, direct job satisfaction consists of four terms:

Absolute individual performance (y_i).

The absolute performance of the teammates, though a direct effect ($b\bar{y}_{-i}$), because the agent may care directly about the team output.

Relative performance ($\rho(y_i - \bar{y}_{-i})$), whose strength depends on how observable postings are. This can be rationalized as a reduced form of Benabou and Tirole's (2003) model of worker type signaling.

Cost of effort ($\frac{1}{2}c_i y_i^2$), reflecting the assumption that high performance requires more work and that certain agents are more skilled (lower c_i). For future reference, let $c_{\min}$ be the cost of the best agent and $c_{\max}$ the cost of the worst agent.

For lean management, we recognize that it constitutes a fundamental change in the relationship between the firm and its workforce. It affects expectations, beliefs, processes, and incentives in multiple ways. However, in the experiment under consideration employees are only involved with the first phase of the "lean journey," which is primarily on communicating the cultural principles of lean management. In particular, employees become familiar with the "Cultural Enabler" concepts of respect and humility. As Toyoda (1950) put it: "Humility is considered the quality of being modest, unassuming in attitude and behavior. It can also be taken as a feeling or showing respect and deference toward other people." The spirit of humility and respect aims to induce employees to shift from a focus on individual outcomes to collective outcomes.

Therefore, we model the change in the relational contract as a modification of the reference points that are used to assess employee success within the organization. Pre-Phase 1, performance was assessed primarily at the individual level: driver by driver. After the change, success includes a team component: site by site. Now, drivers care not just about their job satisfaction but also about those of their teammates (see Sobel 2005 for a survey of interdependent preference models). This can be because of two reasons. First, the new relational contract is a "psychological contract" that changes the culture of the firm, which in turn leads to a change in the preferences of employees with a greater emphasis on the welfare of their colleagues (Rousseau 1995). Second, even if preferences remain stable, an increased emphasis on team performance may lead workers to be more sensitive to the feelings of their teammates, since negative feelings (such as resentment toward a high-performer) may jeopardize future team cooperation.

We capture the effect of the initiation of Phase 1 in the most parsimonious way. The shift from individual job satisfaction to team job satisfaction is represented as an increase in the importance of the reference group, which in this case is the team the worker belongs to. Namely, recalling that u_i is the direct job satisfaction of agent i, we define U_i as the overall job satisfaction of i and we assume that it depends on his own direct satisfaction but also on that of his coworkers:

$$U_i = (1 - \lambda)\, u_i + \lambda \bar{u}_{-i} \ \ ,$$

where: U_i is overall job satisfaction; λ is a parameter that captures the extent to which the collectivist focus has been internalized by employees (with $\lambda = 0$ being pure individualism and $\lambda = 1$ representing absolute selflessness); and $\bar{u}_{-i}$ represents the average direct utility of the other agents, namely

$$\bar{u}_{-i} = \frac{\sum_{j \neq i} u_j}{n - 1} \ \ .$$

The structure of the present model parallels that of the model used in Bandiera, Barankay, and Rasul (2005), where each worker puts some weight on his own payoff and some weight on the payoffs of his or her coworkers.

As mentioned above, the weight λ has a direct interpretation as an other-regarding preference or as an indirect interpretation as a desire to maintain a good team spirit in order to keep the tam productive. In either case, the introduction of Phase 1 results in an increase in λ.

Now that we have a model that encompasses the introduction of performance postings and/or Phase 1 of lean management, we are ready to characterize the effect of the two practices on employee performance:

Proposition 1 *In equilibrium:*

(i) Pre-Phase 1 ($\lambda = 0$), the introduction of postings has a positive effect on agent performance;

(ii) There is a negative complementarity between Phase 1 and postings:

$$\frac{\partial^2 \hat{y}_i}{\partial \lambda \partial \rho} < 0$$

(iii) If the presence of collectivist culture (Phase 1) is sufficiently strong (λ is large), introducing postings worsens agent performance.

(iv) The dispersion of performance across agents displays a negative complementarity between Phase 1 and postings:

$$\frac{\partial^2}{\partial \lambda \partial \rho} \left(\max_i y_i - \min_i y_i \right) < 0$$

Proof. The overall job satisfaction of agent i is given by

$$\begin{aligned} U_i &= (1-\lambda)\,u_i + \lambda \frac{\sum_{j\neq i} u_j}{n-1} \\ &= (1-\lambda)\left(y_i + b\frac{\sum_{j\neq i} y_j}{n-1} + \rho\left(y_i - \frac{\sum_{j\neq i} y_j}{n-1}\right) - \frac{1}{2}c_i y_i^2\right) \\ &\quad + \lambda \frac{\sum_{j\neq i}\left(y_j + b\frac{\sum_{k\neq j} y_k}{n-1} + \rho\left(y_j - \frac{\sum_{k\neq j} y_k}{n-1}\right) - \frac{1}{2}c_i y_j^2\right)}{n-1} \quad . \end{aligned}$$

Hence, the marginal effect of a performance increase on agent i's overall satisfaction is given by

$$\begin{aligned} \frac{dU_i}{dy_i} &= (1-\lambda)(1+\rho - c_i y_i) + \lambda \frac{\sum_{j\neq i}\left(b\frac{1}{n-1} - \rho\frac{1}{n-1}\right)}{n-1}; \\ &= (1-\lambda)(1+\rho - c_i y_i) + \lambda \frac{b-\rho}{n-1}; \end{aligned}$$

yielding first-order condition

$$\hat{y}_i = \frac{1}{c_i}\left((1+\rho) + \frac{\lambda}{1-\lambda}\frac{b-\rho}{(n-1)}\right). \tag{1}$$

Hence

$$\begin{aligned} \frac{\partial \hat{y}_i}{\partial \lambda} &= \frac{1}{c_i}\frac{1}{(1-\lambda)^2}\frac{b-\rho}{n-1}; \\ \frac{\partial \hat{y}_i}{\partial \rho} &= \frac{1}{c_i}\left(\frac{(1-\lambda)(n-1) - \lambda}{(1-\lambda)(n-1)}\right); \end{aligned}$$

and therefore

$$\begin{aligned} \left.\frac{\partial \hat{y}_i}{\partial \rho}\right|_{\lambda=0} &= \frac{1}{c_i} > 0; \\ \frac{\partial^2 \hat{y}_i}{\partial \lambda \partial \rho} &= -\frac{1}{c_i}\frac{1}{(1-\lambda)^2}\frac{1}{n-1} < 0; \end{aligned} \tag{2}$$

which proves (i) and (ii). Also

$$\frac{\partial \hat{y}_i}{\partial \lambda} < 0 \quad \text{if} \quad \rho > b \ ,$$

which proves (iii).

For (iv), note from (1) that y_i can be expressed as

$$\hat{y}_i = \frac{1}{c_i} f(\lambda, \rho) \quad ,$$

where, as we saw in (2), $f(\lambda, \rho)$ exhibits negative complementarities.

The highest performance is by the agent with the lowest cost, $c_{\min}$, and the lowest performance is by the agent with highest cost: $c_{\max}$. Therefore

$$\begin{aligned}\frac{\partial^2}{\partial\lambda\partial\rho}\left(y_{\max}-y_{\min}\right) &= \frac{\partial^2}{\partial\lambda\partial\rho}\left(\frac{1}{c_{\min}}f\left(\lambda,\rho\right)-\frac{1}{c_{\max}}f\left(\lambda,\rho\right)\right) \\ &= \left(\frac{1}{c_{\min}}-\frac{1}{c_{\max}}\right)\frac{\partial^2}{\partial\lambda\partial\rho}f\left(\lambda,\rho\right)<0\end{aligned}$$

■

The main result of the model is point (iii). The effect of introducing relative postings is moderated by the initiation of the cultural transformation of Phase 1. The effect of postings on performance becomes lower as the collectivist culture becomes more pervasive. This is because Phase 1 makes workers more concerned about the "ego bashing" effect on one's team members (and its subsequent effect on relationships) that results from making some individuals' high l performance salient. Thus, as concern about team members and team spirit increases, high performers are less inclined to over-perform and out of concern about causing their teammates to feel badly.

The other results are easy to understand once (ii) is in place. Absent Phase 1, postings improve performance because agents value better performance relative to their colleagues (point i). The negative complementarity between Phase 1 and postings means that if the culture intervention is sufficiently strong the effect of the introduction of postings on performance must be negative (point iii).

Point (iv) is due to the joint effect of more productive agents having more room to reduce performance and more productive agents having the incentive to reduce performance under Phase 1 and performance posting in order not to hurt their less productive team members.[5]

4 The Experiment

The experiment occurred between August 2013 and July 2014 as EOBR was rolled out throughout the company. Beginning in August, 2013, the EOBR system was installed on a weekly basis onto trucks located in a predetermined set of sites. This rollout lasted four months, so that all trucks in the company had received the system by December, 2013. Piggybacking off this rollout, we implemented a three by two research design in which we assigned three performance posting conditions randomly across Phase 1 and pre-Phase 1 sites during this period.

The three posting conditions were stratified by week and Phase 1 status. In other words, for the set of

[5] The dispersion of agent performance is represented in the proposition by the range $\max_i y_i - \min_i y_i$, but it applies to other second-moment measures as well.

sites scheduled for EOBR installation within a given week, we assigned the control, Treatment Group 1 and Treatment Group 2 conditions evenly within Phase 1 and pre-Phase 1 sites. The actual performance postings began six weeks after the EOBR installation, allowing for system calibration and pre-treatment data collection.[6]

4.1 Performance Postings

We designed two posting treatments in addition to the control group: one in which the driver names were posted next to performance information and one in which the employee IDs were used. In this latter treatment, a driver can identify his own standing and view the distribution within the site, but does not know any other individual's performance, nor do others know his performance. We make use of this latter condition in later sections when we provide evidence in support of the underlying mechanism that we propose drives our main result. Because of the substantial number of sites and the lack of pre-existing outcome data to perform power analyses, we placed equal numbers of sites into each of the three conditions (control, named and IDed postings).

The posting were refreshed on a weekly basis, beginning six weeks from the EOBR rollout date for a given site. We stratified the assignment of these postings by week and Phase 1 status. As such, on any given week, an equal number of sites with control, Treatment Group 1 and Treatment Group 2 assignments would be rolled out, in proportion with the Phase 1 and pre-Phase 1 sites that had received the EOBR system six weeks prior.[7] This timing allowed us to obtain thirty days of pre-measures (we discarded the first two weeks while the systems were calibrated to the trucks). The pre-measures, combined with the control group and Phase 1 stratification, enable the triple-differences research design that we describe in Section 5.2. See Appendix Figure A1 for a visual depiction of the timing of the experiment.

The postings contain the employee identifier (either driver names or employee IDs, depending on the treatment assignment) and all four performance metrics recorded by the EOBR system. These metrics are Gap score, Shift score, Excess idle time and Total fuel lost. For completeness, we report all four outcomes in our analysis. We discuss each of these in more depth below in Section 4.3. See Appendix Figure A2 for a sample of the posting.

[6] The following example illustrates the timing and stratification: three Phase 1 and six pre-Phase 1 sites receive the EOBR system in Week 3. Of these nine sites, we randomly assign one Phase 1 and two pre-Phase 1 sites to the control group and to each of the two treatment groups, respectively. The six sites sites assigned to the two treatment groups begin posting in Week 9, six weeks post-installation.

[7] The following example illustrates the timing and stratification: three Phase 1 and six pre-Phase 1 sites receive the EOBR system in Week 3. Of these nine sites, we randomly assign one Phase 1 and two pre-Phase 1 sites to the control group and to each of the two treatment groups. The six sites sites assigned to the two treatment groups begin posting in Week 9, six weeks after the EOBR installation.

4.2 Sites Included in Field Experiment

In discussions with the company senior management on the need for quasi-random assignment of Phase 1, they mentioned that the earliest ten sites were selected specifically to be pilot sites, with various reasons for their inclusion. As a result, we conservatively discarded all sites launched in that first year.[8] We further excluded 36 sites that were scheduled to launch Phase 1 during the timeframe of the study, as these locations could not be reasonably be classified as either Phase 1 or pre-Phase 1. Lastly, we discarded 72 sites scheduled for treatment or corresponding control during the first half of the study. During our mid-project checkpoint, we discovered that no formal verification process of the treatments had been implemented and upon further investigation, we learned that there was marginal compliance up to that point. After this discovery, the company instituted a formal process to verify that performance was posted as required by the experiment guidelines, include weekly photographs of the postings, conference calls, and a shared spreadsheet tracking system. This last exclusion does not introduce any bias since it applies universally to all sites scheduled for treatment or control during the first half of our experiment. After these adjustments, the experimental sample included approximately 5000 unique drivers in 143 sites, 47 in the control group, 50 in the named postings group ("Treatment 1") and 46 in the IDed postings group ("Treatment 2").

To ensure that the routes in our dataset are comparable, we then excluded inter-city routes (defined as routes above 300 miles) and routes with EOBR data that was clearly unreasonable (less than 15 mile routes or MPG <1 or >15, less than 1% of the sample). This left us with a sample of 330,689 driver-days.

Because the company did not have the managerial bandwidth to reinforce the importance of the performance postings on a continual basis, we expected to see some diminishment of the effects of the postings over time. For our main multivariate analyses, therefore, we restricted the windows of the experiment from the thirty days prior to the postings to the thirty days after. We also removed the five days immediately surrounding the scheduled posting dates, since many of the site managers chose a different day of the work week to post performance to coincide with group meetings, rather than on the date specified by the experiment. We were left with a sample of 93,913 driver-days within these narrowed windows that we use in our primary multivariate analyses. In the appendix we repeat the analyses with the full, long windows and show that the results are largely replicated, although with somewhat larger standard errors.

Appendix Tables A1 and A2 contains a summary of the sample construction. Note that the sample of 143 sites used in the experiment is representative of the sites within the firm, based on observable site characteristics and pre-posting driver performance.

[8] The actual number of sites discarded in this step has been masked for confidentiality purposes.

4.3 Outcome Variables

We focus on all four available outcome variables for this study. These measures are designed by the EOBR provider to capture different aspects of efficient driving performance. Gap score calculates the difference between the average actual and "potential" miles per gallon expended on a given route. The potential miles per gallon is calculated by the EOBR system based on what it considers to be optimal shifting and speed patterns, given weather conditions and route characteristics. Gap score is represented in percentage terms such that, for example, if actual and potental mpgs for a given route are 6.5 and 7.0, respectively, the Gap score would be 7.1 ((7.0-6.5)/7.0*100). A higher Gap score, therefore, represents worse (less efficient) driver performance.

Shift score is the percent of shifting events performed on the route that remains within designated revolutions per minute limits for the engine. For example, if a driver shifts five hundred times on a given route, her Shift score will be 90 if she revs the engine above a designated threshold during fifty of those shift events. In order to standardize with the other three outcomes, we reverse-scored the measure for our multivariate analyses, so that a higher value denotes worse shifting performance.

Excess idle time is a measure of the minutes that an engine idles beyond a designated time period, thereby wasting fuel. This metric particularly captures instances in which the driver allows the engine to idle while making a delivery, counter to company policy.

Lastly, Total fuel lost is an aggregate measure of all the fuel wasted from idling, inefficient shifting, speeding and gearing (the latter two factors are unavailable as separate measures). As with Gap score, a higher value for Excess idle or Total fuel lost represents worse performance.

Each measure is intended to be independent from the others, with different behaviors required to improve each of the scores.[9] Because all four measures are included in the weekly postings, we investigate each of them as outcome variables in our analysis.

4.4 Balance of Assignment

Table 1 reports the balance statistics between the control group and the two different treatment groups in the field experiment. Within the sample, the Control and Treatment Group 1 (names) groups are statistically indistinguishable, while the Treatment Group 2 (IDs) group are indistinguishable on most variables, including pre-treatment driver performance. However, sites in Treatment Group 2 are somewhat likelier to be Phase 1 sites. This imbalance was introduced during the treatment assignment process because the authors were supplied with an outdated schedule of the Phase 1 rollout, upon which the stratification relied.[10] Since the results of

[9] With the possible exception of Total fuel lost, although its association with idling and shifting is not straightforward and depends on the make of the engine and other attributes of the truck, all of which are available to the EOBR system.

[10] Once the classifications were updated, the experiment had progressed beyond the point at which a reallocation across treatments was feasible. 13 sites were misclassified, 4 as Phase 1 (in which Phase 1 had not yet launched) and 9 as pre-Phase 1 (in which Phase

our proportional sensitivity analysis suggest that omitted variable bias is not driving Phase 1 assignment, we do not believe that this stratification error affects the interpretation of our findings. Nevertheless, due to this issue, we interpret results concerning the Treatment Group 2 with somewhat more caution than those of the Treatment Group 1. We also conduct the analysis on a matched sample (right hand columns of Table 1) in which the imbalance is eliminated.

Table 2 reports the balance statistics between Phase 1 and pre-Phase 1 sites. Here, we find several observable differences between Phase 1 and pre-Phase 1 sites, primarily in the average site size (as measured by "Tractors per site"). For our experiment, this size difference would present a challenge in interpreting our main result if larger sites are both Phase 1 and likelier to resist performance postings for unrelated reasons. The table also shows other differences between sites, including lower MPG in Phase 1 sites, possibly reflecting more urban locations. We consider these differences in the section below, "Randomness of Phase 1 Assignment." < < < Insert Tables 1 and 2 about here > > >

5 Impact of Postings and Collective Values on Driver Performance

5.1 Preliminary Evidence

Figure 2 depicts the combined response of both Phase 1 and pre-Phase 1 sites to the driver postings.[11] The x-axis is normalized such that week 0 represents the week that performance was posted at each site, regardless of the calendar date of each posting. The y-axis measures the difference in outcome between the treatment groups and the control group, with the 0-line indicating no difference between treatment and control groups, values above zero signifying worse performance and values below 0 signifying better performance.

Three conclusions can be drawn from this figure. First, there are no significant pre-treatment differences between the groups. Second, there is no discernable treatment effect. Lastly, the two treatment groups exhibit similar patterns.

< < Insert Figure 2 about here > >

Figure 3 depicts the response to performance postings by Phase 1 and pre-Phase 1 sites and show graphically one of the main results of the experiment. For visual clarity, we separate the two treatment groups into separate plots.

Three conclusions are apparent from this figure. These plots shows a clear differences in how Phase 1 and pre-Phase 1 sites respond to the named postings, with the pre-Phase 1 sites in Treatment Group 1 showing relatively better performance than the control group (below the 0-line) and the Phase 1 sites showing relatively

1 had already launched at least 6 months prior to the commencement of the experiment).

[11] Figures 2 and 3 show driver response as measured by Gap score. Appendix Figures A4 and A5 replicate this analysis using the other three outcome measures.

worse performance (above the 0-line). Second, the difference between the control and Treatment Group 1 (named posting) appears fairly persistent over the timeframe of the experiment, although there appears to be some convergence near the final weeks of each plot. Two points are important to note regarding this convergence: there is more noise in these final weeks than in the earlier weeks, since only the sites with the earliest EOBR installations had data that extended this far at the time of the analysis. Also, the company did not mandate consistent communication to reinforce the importance of the performance postings throughout the study period and thus we expected some habituation as the postings were refreshed.[12] As a result, it is not possible to infer whether apparent convergence is an artifact of the data and experiment or a more general finding.

Third, we do not see similar patterns in the Treatment Group 2 (IDs). The Phase 1 and pre-Phase 1 sites are statistically indistinguishable from each other and from the control group.

$< <$ Insert Figure 3 about here $> >$

Overall, this figure shows preliminary evidence for one of our primary conclusions: that drivers' responses to the performance postings depend on whether their site had initiated Phase 1. We interpret these results as showing that drivers' reactions to the postings depends on the relational contract of the site to which she/he is assigned.

5.2 Intent to Treat Estimates

We now turn to multivariate analyses. We estimate the differential impact of the postings on Phase 1 and pre-Phase 1 sites using the following triple-differences equation:

$PERF_{it} = \alpha TREAT_i * PHASE1_i * POST_{it} + D'_{it}\beta + e_{it}$ (1) where i represents a given driver and t is the calendar date. $PERF$ is one of the four performance outcomes, $TREAT$ is a vector of two indicator variables, one for each of two possible posting assignments (named or IDed performance postings), and $PHASE1$ is an indicator variable that is equal to one if the sites have launched Phase 1 at least three months before the beginning of the experiment. $POST$ is equal to one after the assigned date of the posting rollout for the two treatment groups or, equivalently, six weeks after the EOBR rollout for the control group. All pair-wise interactions and individual variables associated with the triple-differences term are also included in the model and α represents the vector of coefficient estimates for all the associated terms. We are primarily interested in the coefficient on the triple interaction itself, which estimates the difference in response to the performance postings between the Phase 1 and pre-Phase 1 sites. D_{it} is a vector of control variables that includes the total number of tractors at the site to measure the size of a site, calendar date fixed effects, Phase 1 manager fixed effects, regional fixed effects, and the distance and potential MPG of the route. We cluster standard errors by

[12] Nor, as we have discussed above, did they attach any explicit incentives to the performance results, which we consider an advantage of this study.

site.

We also perform several variations of this analysis to further probe the validity of our initial results. First, we add in driver and date fixed effects to control both for driver traits and seasonality. Second, we create a subsample of the data that matches Phase 1 and pre-Phase 1 sites to account for the quasi-random assignment of Phase 1 in the experiment. Lastly, we do an instrumental variables analysis where we instrument actual postings with assigned postings to account for incomplete compliance. We discuss each of these analyses in turn below.

5.2.1 Combined Effect Across All Sites

We begin by estimating the simple intent-to-treat model without differentiating between Phase 1 and pre-Phase 1 sites. Table 3 shows the results of all four performance outcomes, both without and with controls (odd and even columns, respectively). Consistent with Figure 2, the posting intervention appears to have no effect. Without accounting for the underlying relational contract at the site, therefore, we might inaccurately conclude that performance postings have no effect on worker performance.

$<<$ Insert Table 3 about here $>>$

Table A3 replicates the analysis on the matched sample, with similar results.

5.2.2 Effect by Phase 1 and Pre-Phase 1 Sites

We next estimate the differential impact between Phase 1 and pre-Phase 1 sites.Table 4 shows the results of the intent-to-treat analysis modeled in Equation (1) and is consistent with the plots in Figure 3. In Columns (1) and (2), we see a large, positive difference in the response of Phase 1 sites to the named posting treatment, where positive differences correspond to worse performance. Using the estimates in Column (2), we observe a 13.6% greater average Gap Score within Phase 1 sites with named postings (coefficient on *Post**Treatment Group *1*Phase 1)*, relative to pre-Phase 1 sites, and a 4.0% lower average Gap score within pre-Phase 1 sites with named postings relative to control (*Post**Treatment Group *1,* albeit insignificant).[13]

No similar effect was estimated for the second treatment group that posted the IDed performance, with *Post**Treatment Group *2*Phase 1* insignificant across all specifications. The estimated coefficients for the other pairwise interactions and single variables correspond to our expectations. Consistent with successful randomization, we estimate no statistical difference between treatment groups and control groups (coefficients on *Treatment Group 1* and *2* and Treatment Group *1*Phase 1* and Treatment Group *2*), nor any difference between the control groups in the post-period (coefficient on *Post*Phase 1*). Notably, the deterioration of the

[13] Since we have a randomized experiment, our results should hold regardles of pre-period data. We show this result in Appendix Table A4, which replicates Table 4 using post-period data only.

driving performance within Phase 1 sites in Treatment Group 1 (named posting) is generally more significant and larger than the improvement in the pre-Phase 1 sites. For example, the 5.5% increase in Total fuel lost (Column 8) for Phase 1 sites relative to pre-Phase 1 sites corresponds to a 3.8% deterioration relative to control versus the 1.7% improvement estimated for the pre-Phase 1 sites.

Also notable here and in subsequent analyses is that the coefficient estimates are stable with and without the inclusion of control variables, consistent with successful randomization of both the posting and Phase 1 assignment. We explore the assumption of random Phase 1 assignment explicitly in the next section.

< < Insert Table 4 about here > >

Table 5 repeats the analysis including driver fixed effects and the results are similar.[14] As in Table 4, we observe no underlying pattern for Treatment Group 2, the IDed posting group. The magnitude of the coefficients range from nearly identical (for Shift score) to 30% lower (for Idle time), indicating that the effect of performance postings is due mostly to changes in driver behavior, rather than compositional differences between sites.[15]

< < Insert Table 5 about here > >

5.3 Randomness of Phase 1 Assignment

One challenge of the study is that Phase 1 status was not assigned via experimental randomization. A mitigation of this concern is that the company's management indicated to us that, after the 2011 pilot period (not included our sample), the choice to launch Phase 1 at a given location was driven by easing the travel burden on the regionally-based managers involved in the initiative, without consideration of the likely success of the initiative at a given site. When asked whether the rollout was related factors that might affect the acceptance of performance postings, the manager in charge of Phase 1 explicitly told the study authors that this was not the case.

However, as is evident from the balance analysis in Table 2, this rollout strategy still resulted in observable differences between sites between Phase 1 and pre-Phase 1 sites, particularly size and average MPG, as discussed in Section 4.4.

We address this imbalance in two ways. First, we attempt to characterize the magnitude of omitted variable bias required to explain our treatment effect and assess whether this is reasonable. We perform a proportional sensitivity analysis as developed by Oster (2014), based on Altonji, Elder and Taber (2005), for this analysis. Second, we reproduce our main analysis on a propensity-score matched sample in which Phase 1 and pre-Phase 1 sites are balanced (see Tables 1 and 2 right hand columns).

[14] One difference between these tables is that we see a large improvement in Shift scores for the control group for Shift after the treatment. This result is not replicated in our other tables, including in our matched analysis with driver fixed effects (Table 8) or in our long window with driver fixed effects (Tables A10 and A11). We treat this result as an anomaly.

[15] The differences in these coefficient magnitudes between specifications with and without driver fixed effects are smaller than those found by Lazear (2000). In that study, roughly 50% of the productivity improvement was due to compositional effects, whereas the upper bound for our study is 30 percent.

5.3.1 Estimating Magnitude of Potential Omitted Variable Bias

For this analysis, we appeal to Oster's (2014) method of estimating omitted variable bias based on treatment coefficient stability and the incremental explained variance of the included controls. The central assumption behind this method is that the ratio of coefficient movements to explained variance of omitted variables is proportional to that same ratio of observable controls. This proportionality, δ, can be estimated as long as the upper bound on explainable variance is specified ($R_{\max}$=Π*$\tilde{R}$). Here, $\tilde{R}$ is the explained variance of the baseline specification with full controls, $R_{\max}$is the maximum explainable variance and Π is the multiplier factor between the two measures. The logic is that any estimate of $|\delta|$ that is greater than 1 assumes that the contribution of the omitted variables is more than the contribution of the observed controls. We follow Oster's (2014) suggested standard: to report δ for a given Π , where Π is suggested to be 1.3, a threshold judged to be reasonable based on prior randomized studies. We also report a substantially stricter threshold of $\Pi = 2$ to assess the sensitivity of our analysis to these parameters. The test can also report the estimated value of β for $\delta = 1$ as an alternative reporting statistic. For completeness, we report both δ and β.

Table 6 shows the results of this test. Because the test is designed to be applied to a simple treatment model, we reduce our baseline triple-differences model to eliminate the interaction variables, many of which included the *Phase 1* treatment variable. We do this in two ways. First, in Panel A, we drop the pre-treatment observations to reduce the triple-difference model to a diffs-in-diffs. We also drop all observations in Treatment Group 2 (the ID-ed Treatment Group), since our primary result is in Treatment Group 1 (the named group). In this specification, the treatment effect is Treatment Group *1*Phase 1,* with Treatment Group *1* used as a control. As an alternative approach, in Panel B we further reduce the diffs-in-diffs model to a simple treatment model by dropping the control groups. In this specification, our treatment variable is *Phase 1.* This model is the simplest; however, it understates the treatment effect of Phase 1, since there are offsetting differences in the control groups that are unaccounted for in this specification.

Regardless of which of these two approaches are used, we can see from Table 6 that, according to this method, omitted variable bias in the Phase 1 assignment does not appear to be driving our treatment results. Using a Π value of 1.3, our estimates of $|\delta|$ range from 3.0 to 18.0, well above the threshold value of 1.0. Even using the much stricter threshold value of $\Pi = 2.0$ our estimated values of $|\delta|$ all exceed 1.0, most by substantial margins. Relatedly, the adjusted treatment values of β^* for δ fixed to 1 are all in the range of our estimated treatment effects.

$<<$ Insert Table 6 about here $>>$

5.3.2 Propensity Score Matched Analysis

We supplement the above analysis by constructing a matched subsample of sites that corrects for the imbalance between Phase 1 and pre-Phase 1 sites. The matched sample was created using a propensity-score approach that matches Phase 1 and pre-Phase 1 sites on size (the primary observable metric by which the Phase 1 and pre-Phase 1 sites differ), average trip distance (a proxy for proximity to urban market), and pre-treatment measures of MPG, Gap score and Shift score. The matched sample includes 82 of the 143 sites in the full experimental sample. The excluded sites include seven of the larger, urban sites in the Phase 1 group and fifty four of the smaller sites in the pre-Phase 1 group.

The right side of Tables 1 and 2 compares the sites across these two groups in this reduced sample and shows that the sites are now statistically indistinguishable across both observable site characteristics and pre-posting driver performance.

The right hand side columns of Appendix Table A2 display the descriptive statistics of the matched sample and allows comparison both to the all the sites at the company and to those sites included in our full sample.

Table 7 reproduces the ITT estimates of Table 3 on the matched sample. The point estimates generally increase and also represent slightly larger percentage standard deviation increases (although they are not statistically different from each other). For example, the Column (6) estimate on *Post*Treatment Group 1*Phase 1* of 0.0482 represents a 23% standard deviation increase in Log Idle time compared to 17% increase based on the estimates in Table 4 Column (6). In general, the results of Table 3 are reproduced and, if anything, strengthened, with the exception of Shift score.

< < Insert Table 7 about here > >

Table 8 replicates the driver fixed effects analysis of Table 5 on the matched subsample. The results are largely reproduced, again with the exception of Shift score, and show similar 0-30% decrease in coefficient estimates relative to estimates without driver fixed effects.

< < Insert Table 8 about here > >

5.4 Additional Analyses

5.4.1 Potential Changes in Underlying Route Characteristics

We also performed three additional analyses to further rule out potential concerns about our data and experimental design. One such potential concern is whether there are fundamental differences in the or characteristics of the routes driven in Phase 1 and pre-Phase 1 sites that may explain our observed effects. While we believe that this possibility is remote - it would have to affect only Treatment Group 1 (named posting), only at the same six week post-EOBR rollout window as the performance postings and also be orthogonal to the site characteristics on which we based the matched analysis - we perform a placebo test to further rule out this possibility. For this test, we replace our four outcome variables with "potential" MPG. Potential MPG is the system-calculated variable that response to route characteristics and road and weather conditions, but not to driver performance. Therefore, if any route characteristics changed during this period in the Phase 1 named-posting group that led to changes in driver performance, we should observe similar patterns in the potential MPG metric.

The results of this analysis are shown in Appendix Tables A5 on the full sample and A6 on the matched sample. Potential MPG shows no changes during this period, while Actual MPG - which is directly related to driver effort - does. It does not appear, therefore, that underlying changes to the routes are driving the results.

5.4.2 Correcting for Compliance

Coordinating the posting rollouts posed a management challenge for the company, particularly since the postings were rolled out on a weekly basis across 48 states during the busy winter season. As a consequence, approximately 65% of the sites in our experimental sample fully complied with the postings, while the remainder either did not comply or only partially complied. The sites with full compliance were statistically indistinguishable from the other sites in terms of Phase 1 status, size, treatment group and other observable attributes. To account for this incomplete compliance, we instrumented actual treatment with assigned treatment. The results of this analysis are shown in Appendix Tables A7 and A8 and are stronger than our earlier analyses (including fixed effects and the matched cohorts).

5.4.3 Persistence of Effect

Finally, Appendix Tables A9-A12 repeat the analysis shown in Tables 4, 5, 7 and 8 without restricting the time windows to the 30 days pre- and post- performance postings. These analyses now include 47 days prior and 207 days after the performance postings. We find that, consistent with some attenuation, the standard errors of the estimates are generally larger, but the effect sizes are close to the narrower-window analysis. As we discuss in Section 5.1, since the company did not consistently reinforce the postings during our experiment, we expected some reversion in behavior but still see an effect beyond 30 days.

6 How do We Know it is "Collectivistic Orientation" that Matters?

Up to this point, we have reasoned that the Phase 1 intervention created a collectivist-oriented relational contract and that it is this collectivist orientation that drives the difference in employee response to performance postings. However, since relational contracts are, by definition, extremely hard to observe, how do we know that this is the mechanism at work in this experiment?

The nature of relational contracts makes it difficult to answer this question definitively. However, in this section, we present three distinct tests that, taken together, are consistent with our argument. In our first test, we examine the differences in the response between our two treatment groups, the first of which identified the driver performance by name and the second of which identified performance by the anonymous employee ID. In the second test, we look at the effect of postings on the dispersion of performance within each location. Lastly, we relate driver performance to a proxy measure of "collectivist orientation" based on an employee engagement survey. Each of these tests draws on insights from social psychology to construct predictions that test the role of collectivist orientation. The results of these tests strongly support our proposed mechanism and its emphasis on the role of collectivist-oriented relational contracts in explaining the effects of Phase 1.

6.1 Named vs IDed Postings

Our reasoning suggests that the differential effect of the feedback intervention, in both Phase 1 and pre-Phase 1 sites, is driven by the identification of individuals on the performance postings. Having anonymous rather than named postings should undermine the mechanism that we propose for the negative effects of performance posting among our Phase 1 sites, where a collectivistic orientation prevails. That is, since social comparisons to specific, known peers are not possible when performance is anonymous, they are less likely to lead to an adversarial dynamic among teammates. Relatedly, anonymous postings should also cause less embarrassment for low performers, and therefore high performers' concerns about harming their teammates' satisfaction are less likely. We would expect, then, that concerns about teammates' feeling and damaged social relations in Phase 1 sites will be attenuated when it is unknown who is beating whom, who is disappointing whom, etc.

Overall, anonymity should reduce the competitive nature of performance postings, removing the element central to the positive and negative effects of posting on performance, respectively, in our pre-Phase 1 and Phase 1 sites. If correct, then anonymous postings should not replicate the pattern of named posting effects that we present above. If incorrect—for instance, if named postings have their effect simply because they convey relative performance feedback (Lazear and Rosen 1981) or even more simply because they convey individual performance feedback—then we would expect that anonymous postings would demonstrate the same pattern of effects as we found for named postings.[16]

[16] This is because individual competition, which we propose motivates drivers in pre-Phase 1 sites where an individualistic orientation prevails, will be stronger in situations in which one's competitors are known, such as cases in which the postings clearly identify where each individual stands, who specifically is beating whom, and who specifically one needs to outdo in order to achieve a higher (Bendersky and Hays 2012; Maholtra 2010). Similarly, the goal of basking in recognition from others and avoiding the embarrassment of being revealed as a poor performer are only relevant in cases where the postings personally identify each individual (Garcia, Tor and Gonzalez 2006). If postings were instead anonymized —where individuals could only identify their own performance —then the incentive of positively distinguishing oneself (which motivates those with an individualistic orientation) would dissipate. More generally, a lack of identifiable social comparisons would diminish the relational component of competition, a critical component of competition among individuals (Garcia and Tor 2009; Johnson and Johnson 1999; Kilduff, Elfenbein, and Staw 2010).

We utilized this distinction between identifiable and anonymous postings as a means of exploring the validity of our proposed mechanism. In particular, in our study we included our second treatment group—ID postings—which was likewise included in our random assignment of sites to posting treatment. In this additional treatment condition, performance was posted in an identical manner to that utilized in our named posting condition except for one difference: rather than identifying employees by name, we identified them on the postings by their employee IDs.

Consistent with the reasoning above, we find no difference between Treatment Group 2 and the control in our triple-difference analysis so far presented.[17]

6.2 Performance Variance

Another test for our proposed mechanism is to examine performance variance, rather than averages. If the difference in driver response to the postings results from a collective relational contract, we predict that named postings should reduce variance in driver performance within Phase 1 sites and raise variance within pre-Phase 1 sites.

This prediction corresponds to point (iv) of Proposition 1. Intuitively, in a Benabou and Tirole (2003) world, the initiation of Phase 1 will induce top performers to reduce their effort to avoid hurting their teammates' egos. This prediction resonates with a number of findings in the psychology literature. Consistency in reactions will likely be more common in cases that involve a direct violation to the group's values and norms (Branscombe et al. 1999; Spears et. al. 1997; Tajfel and Turner, 1979), as in the case when named postings are introduced in Phase 1 sites. This is because these violations are both salient and threatening, thus heightening identification with the group (Ashforth and Mael 1992; Tajfel and Turner 1979) which in turn leads to greater conformity and more normative response patterns (Blader and Tyler 2009; Haslam 2004).

In sum, our speculation that the effects of Phase 1 are driven by collective orientation would be supported if performance variance among Phase 1/named posting sites is lower than that of other sites. That is, our proposed mechanism would suggest that there should be greater homogeneity of behavior in Phase 1/named-posting sites. In pre-Phase 1 sites, individualistic orientations predominate and thus shared understandings about norms and behavior are diminished. Moreover, in Phase 1/control sites, there is no threat present to draw group members inward to the group, its norms, nor are there factors that would lead high-performers to converge their performance towards the levels of their lower-performing counterparts. Under these circumstances, greater variation among individuals is likely.

In contrast, if performance variance is the same between the Phase 1/named condition and our other conditions, that would call into question our proposed mechanism. In particular, it could either suggest that a)

[17] Tables 4, 5, 7, 8, A5-A12

Phase 1 does not breed a collective orientation and/or b) that the postings are not interpreted by employees as a violation and threat to the prevailing team-based relational contract at Phase 1 sites.

Table 9 shows the effect of performance postings on the coefficient of variation of daily performance across Phase 1 and pre-Phase 1 sites. Several results are apparent. We find that, overall, the coefficient of variation decreases over time (the coefficient on *Post*). This trend may be due to a learning effect on the part of the drivers or improved instrument calibration. Second, we note that this decreased coefficient of variation does not occur in pre-Phase 1 sites with named postings; that is, relative to the control groups, it actually increases in pre-Phase 1 sites with named postings (the coefficient on *Post**Treatment Group *1).* In contrast, in the Phase 1 sites, the coefficient of variation reduces between the control and Treatment Group 1 (named postings) (the coefficient on *Post**Treatment Group *1*Phase 1*).

Lastly, we also observe no effect of IDed postings on variance in either the Phase 1 or the pre-Phase 1 sites. In terms of economic magnitudes, the difference in response between Phase 1 and pre-Phase 1 sites ranges from 25% of a standard deviation of Log(Gap score) variance to 45% of a standard deviation for Log(Idle time). In sum, these results are consistent with our reasoning that collective orientation will compress performance once performance is revealed.

< < Insert Table 9 about here > >

6.3 Engagement Survey Responses

Our third test of the proposed mechanism uses responses to the company's annual engagement survey. This survey was conducted across 45 sites in July 2014. The survey yielded 564 driver responses, approximately a 30% response rate.[18]

From the survey responses, we assemble two indices, each measured at the level of the individual driver. Our first index is a direct measure of collectivistic orientation, based on research in social psychology on group identification. Specifically, we calculate the average responses to questions that assess the degree of teamwork in the workplace, levels of trust and pride in the organization and whether employees feel valued. Our second index is a comparison measure of "instrumental" job satisfaction, similarly based on social psychology, that we use to rule out the possibility that any effects we may find are not specific to collectivist orientation per se, but instead reflect a halo effect of overall employee happiness. This measure is defined as the average responses

[18] One concern with this survey is that it was conducted after our performance data and therefore employee responses may somehow have been affected by the experiment. We use this 2014 survey because the company's 2013 survey was not usable (it was fully anonymized and only aggregated indices per site were available). Despite this timing, a reverse-causal interpretation of our analysis is unlikely: one would have to believe that the weekly postings affected not only driving behavior but also employee attitudes about the company more than four months after they have been removed. A separate concern is the number of sites of the survey: given the costs of using an outside survey provider to run an identified, but anonymized (to the company), survey, the survey was limited to 45 sites. These sites do not provide us with sufficient balance between the six interventions cells ([Phase1 / pre-Phase 1] X [Control / Treatment Group 1 / Treatment Group 2]) to include the main experimental dummies in our analysis.

to questions assessing a driver's belief in the company's future and happiness with his/her compensation and benefits. We compare the effects of these two measures to test our prediction that Phase 1 influences drivers' collectivistic orientation, and that it is this orientation - and not overall satisfaction - in turn that affects drivers' responses to postings.

If collectivist orientation does, in fact, underpin drivers' differential responses to the performance postings, then we should find the following results with the survey data: a) Phase 1 sites score higher on measures of collectivist orientation, and b) collectivist orientation, but not instrumental satisfaction, produces a similar pattern of results as our Phase 1 indicator in our primary triple-differences analysis.

Table A13 shows the relationships between Phase 1 sites and our collective and instrumental indices. This table shows that Phase 1 sites are associated with higher scores across both measures.

Since this is a cross-sectional analysis, we may be concerned that underlying differences between Phase 1 and pre-Phase 1 sites drive the differences engagement survey responses. Accordingly, Table A14 shows the same analysis as the previous table using a sub-sample of survey responses in which Phase 1 and pre-Phase 1 sites were matched by size, region, and our three available demographic measures: race, age and tenure. The results attenuate somewhat but are statistically the same as the unmatched sample. In particular, the association between Phase 1 and collective orientation remains robust.

Table 10 shows the main result of this analysis: a higher score on the collectivist index is associated with a more negative response to named postings, while no such response is observed using the instrumental index. Figure 4 depicts this result graphically. We divide the collective and instrumental indices into deciles and plot each decile's response to the named postings, relative to the lowest decile. We can see from this figure an increasingly negative response to named posting as the collective decile increases, while no such pattern is observed for across the instrumental deciles.

< < Insert Table 10 about here > >

< < Insert Figure 4 about here > >

Table 11 repeats this analysis with driver fixed effects. The results from Table 10 are replicated, with a 10% attenuation of the coefficient estimates.

One cautionary note about this analysis is that there appears to be some pre-treatment differences between drivers with high collective orientation in Treatment Group 1 and the control group (see the *Treatment Group 1*[Category]* coefficient in Table 10), so we cannot rule out absolutely that our observed patterns are not driven by underlying differences between these two groups. Table A15 mitigates this concern by showing that the magnitude of the treatment response appears unrelated to the degree of pre-treatment differences between these two groups.

From this last test, our results produce: a) a pattern for the collective orientation index that is consistent with our proposed mechanism, b) a pattern for the instrumental index that does not explain our primary findings, and therefore c) suggestive evidence that it is not something common to both indices—e.g., a halo effect reflecting generalized happiness—that drives our results but rather something unique to collective orientation.[19]

The three preceding tests support the explanation that Phase 1 affects collectivist orientation, which in turn influence the drivers' responses to the postings. Notably, the tests are independent of each other and yet converge in their support for a collective orientation relational contract. Any alternate proposed mechanism would need to explain not only our main results but also these three distinct sets of results.

7 Conclusion

In this study, we randomized the posting of employee performance postings across a company that was midway through a costly, multi-year process of altering its relational contract with its employees. Employees working in locations with the original contract responded positively to the performance postings, with their performance improving 2-4% relative to the control group (depending on the performance measure). In contrast, employees in the sites with the newer contract responded negatively, with their performance declining 4-13% relative to the control group.

It appears that these different responses are driven by the individualistic orientation of the initial contract and the collectivist orientation of the new contract. This new contract is based on the Toyota Production System which emphasizes the value of teamwork and cooperation, as well as the subservient role of management whose primary task is to enable the front line workers. This result can be understood within an extension of Benabou and Tirole's (2003) model of incentives and prosocial behavior. Research in social psychology and organizational behavior has found that employees respond poorly to perceived inconsistencies in leaders' and organizations' messages. Our findings support this result, with the posting of individualistic performance postings representing a violation of the collectivist contract rolled out by the company. And in fact, presented with the results of this experiment, the company management chose to discontinue the performance postings immediately, concerned about the risk to the new culture they were seeking to instill.

The main contribution of the present paper is to show that the success or failure of a management practice depends on underlying conditions at the firm. These conditions include not just the environment in which the firm operates in and the presence of other management practices, but also on the type of long-term relationship that the firm chooses to establish with its employees. A company who is considering adopting a new practice should ask itself not just whether this practice worked in similar firms, but more specifically whether this practice worked in firms that have a similar relational contract with their employees. This result highlights the

[19] We also created a second comparison measure of "individual job satisfaction" that augmented the instrumental measure with overall job satisfaction and found substantively identical results as with the instrumental measure.

importance of measuring not just management practices but also how workers perceive the relationship with their employer.

We have several directions for future research. In later studies, we aim to randomize the rollout of the relational contract itself, a much lengthier and complex process that should allow us to make more definitive statements about the direct impact of relational contracts on employee productivity. Aside from this direct effect, a further area to explore is the process of altering these relational contracts themselves. Specifically, we would like to understand the factors that determine differences in adoption success. Finally, beyond single firm studies, we would like to extend this research across firms, industries and geographies.

References

[1] Altonji, Joseph G., Todd E. Elder, and Christopher R. Taber, "Selection on Observed and Unobserved Variables: Assessing the Effectiveness of Catholic Schools," *Journal of Political Economy*, 113 (2005), 151-184.

[2] Anderson, Cameron, and Courtney E. Brown, "The Functions and Dysfunctions of Hierarchy," *Research in Organizational Behavior*, 30 (2010), 55-89.

[3] Ashraf, Nava, Oriana Bandiera, and Scott S. Lee, "Awards Unbundled: Evidence from a Natural Field Experiment," *Journal of Economic Behavior and Organization*, 100 (2014), 44-63.

[4] Baker, George, Robert Gibbons, and Kevin Murphy, "Relational Contracts and the Theory of the Firm," *The Quarterly Journal of Economics*, 117 (2002), 39-84.

Baker, George and Thomas N Hubbard. "Make Versus Buy In Trucking: Asset Ownership, Job Design, and Information," *American Economic Review*, 93(2003), 551-572.

[5] Bandiera, Oriana, Iwan Barankay, and Imran Rasul, "Social Incentives in the Workplace," *The Review of Economic Studies*, 77 (2010), 417-458.

———, "Team Incentives: Evidence from a Firm Level Experiment," *Journal of the European Economic Association*, 11 (2013), 1079-1114.

[6] Barankay, Iwan, "Rank Incentives: Evidence from a Randomized Workplace Experiment," *Management Science*, under revision, 2012.

[7] Benabou, Roland, and Jean Tirole, "Intrinsic and Extrinsic Motivation," *The Review of Economic Studies*, 70 (2003), 489-520.

Bendersky, Corinne, and Nicholas A. Hays, “Status Conflict in Groups,” *Organization Science*, 23 (2012), 323-340.

[8] Blader, Steven L., and Tom R. Tyler, "Testing and Extending the Group Engagement Model: Linkages Between Social Identity, Procedural Justice, Economic Outcomes, and Extrarole Behavior," *Journal of Applied Psychology*, 94 (2009), 445-464.

[9] Blanes i Vidal, Jordi, and Mareike Nossol, "Tournaments Without Prizes: Evidence from Personnel Records," *Management Science*, 57 (2011), 1721-1736.

Bloom, Nicholas, and John Van Reenen, “Measuring and Explaining Management Practices Across Firms and Countries,” *The Quarterly Journal of Economics*, 122 (2007), 1351-1408.

[10] ———, "Why Do Management Practices Differ Across Firms and Countries?," *The Journal of Economic Perspectives*, 24 (2010) 203-224.

[11] Branscombe, Nyla R., Naomi Ellemers, Russell Spears, and Bertjan Doosje, “The Context and Content of Social Identity Threat,” in *Social Identity: Context, Commitment, Content*, Naomi Ellemers, Russell Spears, and Bertjan Doosje, eds. Oxford: England, 35-58.

[12] Brown, Rupert J., and John C. Turner, “Interpersonal and Intergroup Behavior,” in *Intergroup Behaviour*, John C. Turner, and Howard Giles, eds. Oxford: Blackwell, 33-65.

[13] Brown, Rupert, and Dominic Abrams, "The Effects of Intergroup Similarity and Goal Interdependence on Intergroup Attitudes and Task Performance," *Journal of Experimental Social Psychology*, 22 (1986), 78-92.

[14] Brynjolfsson, Erik, and Paul Milgrom, "Complementarity in Organizations," in *The Handbook of Organizational Economics*, Robert Gibbons and John Roberts eds. Princeton: NJ, 11-55.

[15] Bursztyn, Leonardo and Robert Jensen, "How Does Peer Pressure Affect Educational Investments?," *Quarterly Journal of Economics*, forthcoming, 2015.

Cowen, Tyler, *Average is Over: Powering America Beyond the Age of the Great Stagnation* (Penguin, 2013).

Josse Delfgaauw, Robert Dur, Arjan Non and Willem Verbeke, "The Effects of Prize Spread and Noise in Elimination Tournaments: A Natural Field Experiment," *Journal of Labor Economics* 33, No. 3 (Part 1, July 2015), 521-569.

Ellemers, Naomi, Russell Spears, and Bertjan Doosje, "Sticking Together or Falling Apart: In-group Identification as a Psychological Determinant of Group Commitment Versus Individual Mobility," *Journal of Personality and Social Psychology*, 72 (1997), 617-626.

Garcia, Stephen M., Avishalom Tor, and Richard Gonzalez, "Ranks and Rivals: A Theory of Competition," *Personality and Social Psychology Bulletin*, 32 (2006), 970-982.

[16] Garcia, Stephen M., and Avishalom Tor, "The N-effect More Competitors, Less Competition," *Psychological Science*, 20 (2009), 871-877.

Gibbons, Robert and Rebecca Henderson, "What Do Managers Do?" in *The Handbook of Organizational Economics,* Robert Gibbons and John Roberts eds. Princeton: NJ, 680-731.

[17] Gruenfeld, Deborah H., and Larissa Z. Tiedens, "Organizational Preferences and Their Consequences," *Handbook of Social Psychology*, 3 (2010), 1252-1287.

Halevy, Nir, Eileen Y. Chou, Adam D. Galinsky, and J. Keith Murnighan, "When Hierarchy Wins Evidence from the National Basketball Association," *Social Psychological and Personality Science*, 3 (2012), 398-406.

[18] Haran, Uriel, and Ilana Ritov, "Know Who You're Up Against: Counterpart Identifiability Enhances Competitive Behavior," *Journal of Experimental Social Psychology*, 54 (2014), 115-121.

[19] Haslam, S. Alexander, *Psychology in Organizations* (Sage, 2004).

[20] Helper, Susan, and Rebecca Henderson, "Management Practices, Relational Contracts, and the Decline of General Motors," *Journal of Economic Perspectives*, 28 (2014), 49-72.

Holweg, Matthias, "The Genealogy of Lean Production," *Journal of Operations Management*, 25 (2007), 420-437.

Ichniowski, Casey, Kathryn Shaw, and Giovanna Prennushi, "The Effects of Human Resource Management Practices on Productivity: A Study of Steel Finishing Lines," *American Economic Review,* 87 (1997), 291-313.

[21] Ichniowski, Casey, and Kathryn Shaw, "Beyond Incentive Pay: Insiders' Estimates of the Value of Complementary Human Resource Management Practices," *Journal of Economic Perspectives*, 17 (2003), 155-180.

[22] Johnson, David W., and Roger T. Johnson, *Cooperation and Competition: Theory and Research* (Edina, MN: Interaction Book Company, 1989).

Kandel, Eugene, and Edward P. Lazear, "Peer Pressure and Partnerships," *Journal of Political Economy*, 100 (1992), 801-817.

[23] Kilduff, Gavin J., Hillary Anger Elfenbein, and Barry M. Staw, "The Psychology of Rivalry: A Relationally Dependent Analysis of Competition," *Academy of Management Journal*, 53 (2010), 943-969.

Koeth, Denise. Mandated EOBRs/ELDs. *Fleet Equipment Magazine*, July 2013.

Lawler, Edward, Susan Mohrman, and Gerald E. Ledford Jr., *Creating High Performance Organizations: Practices and Results of Employee Involvement and TQM in Fortune 1000 Companies* (San Francisco, CA: Jossey-Bass, 1995).

Lawler, Edward, Susan Mohrman, and George Benson, *Organizing for High Performance: Employee Involvement, TQM, Reengineering, and Knowledge Management in the Fortune 1000* (San Francisco, CA: Jossey-Bass, 2001).

[24] Lawler, Edward and Susan Mohrman, *Creating a Strategic Human Resources Organization: An Assessment of Trends and New Directions* (Stanford University Press, 2003).

[25] Lazear, Edward P., "Performance Pay and Productivity," *American Economic Review,* 90 (2000): 1346-1361.

[26] Lazear, Edward P., and Sherwin Rosen, "Rank-Order Tournaments as Optimum Labor Contracts," *The Journal of Political Economy*, 89 (1981), 841-864.

Lemieux, Thomas, William B. MacLeod, and Daniel Parent, "Performance Pay and Wage Inequality," *The Quarterly Journal of Economics*, 124 (2009), 1-49.

[27] Levin, Jonathan, "Relational Incentive Contracts," *The American Economic Review,* 93 (2003), 835-857.

[28] MacLeod, William B., and James M. Malcomson, "Implicit Contracts, Incentive Compatibility, and Involuntary Unemployment," *Econometrica*, 57 (1989), 447-480.

[29] Mael, Fred, and Blake E. Ashforth, "Alumni and Their Alma Mater: A Partial Test of the Reformulated Model of Organizational Identification," *Journal of Organizational Behavior*, 13 (1992), 103-123.

[30] Malhotra, Deepak, "The Desire to Win: The Effects of Competitive Arousal on Motivation and Behavior," *Organizational Behavior and Human Decision Processes,* 111 (2010), 139-146.

Milgrom, Paul, and John Roberts, "The Economics of Modern Manufacturing: Technology, Strategy, and Organization," *The American Economic Review*, 80 (1990), 511-528.

[31] ———, "Complementarities and Fit Strategy, Structure, and Organizational Change in Manufacturing," *Journal of Accounting and Economics*, 19 (1995), 179-208.

Oster, Emily, "Unobservable Selection and Coefficient Stability: Theory and Validation," *NBER Working Paper* No. w19054, 2013 (updated November 2014).

[32] Sobel, Joel, "Interdependent Preferences and Reciprocity," *Journal of Economic Literature*, 43 (2005), 392-436.

[33] Spears, Russell, Bertjan Doosje, and Naomi Ellemers, "Self-stereotyping in the Face of Threats to Group Status and Distinctiveness: The Role of Group Identification," *Personality and Social Psychology Bulletin*, 23 (1997), 538-553.

[34] Stanne, Mary Beth, David W. Johnson, and Roger T. Johnson, "Does Competition Enhance or Inhibit Motor Performance: A Meta-Analysis," *Psychological Bulletin*, 125 (1999), 133-154.

Syverson, Chad, "What Determines Productivity?," *Journal of Economic Literature*, 49 (2011), 326-365.

[35] Tajfel, Henri, and John C. Turner, "An Integrative Theory of Intergroup Conflict," in *The Social Psychology of Intergroup Relations*, William G. Austin and Stephen Worchel, eds. Monterey: California, 33-47.

[36] ———, "The Social Identity Theory of Intergroup Behavior," in *Psychology of Intergroup Relations*, Stephen Worchel and William G. Austin, eds. Chicago: Illinois, 7-24.

[37] Turner, John C., "Towards a Cognitive Redefinition of the Social Group," in *Social Identity and Intergroup Relations*, Henri Tajfel, ed. Cambridge: England, 15-40.

[38] Turner, John C., "Social Categorization and the Self-Concept: A Social Cognitive Theory of Group Behavior," *Advances in Group Processes*, 2 (1985), 77-122.

Womack, Jones, and Daniel T. Jones and Daniel Roos, *The Machine That Changed the World* (Free Press, 1990).

Figure 1: Timeline of Experiment

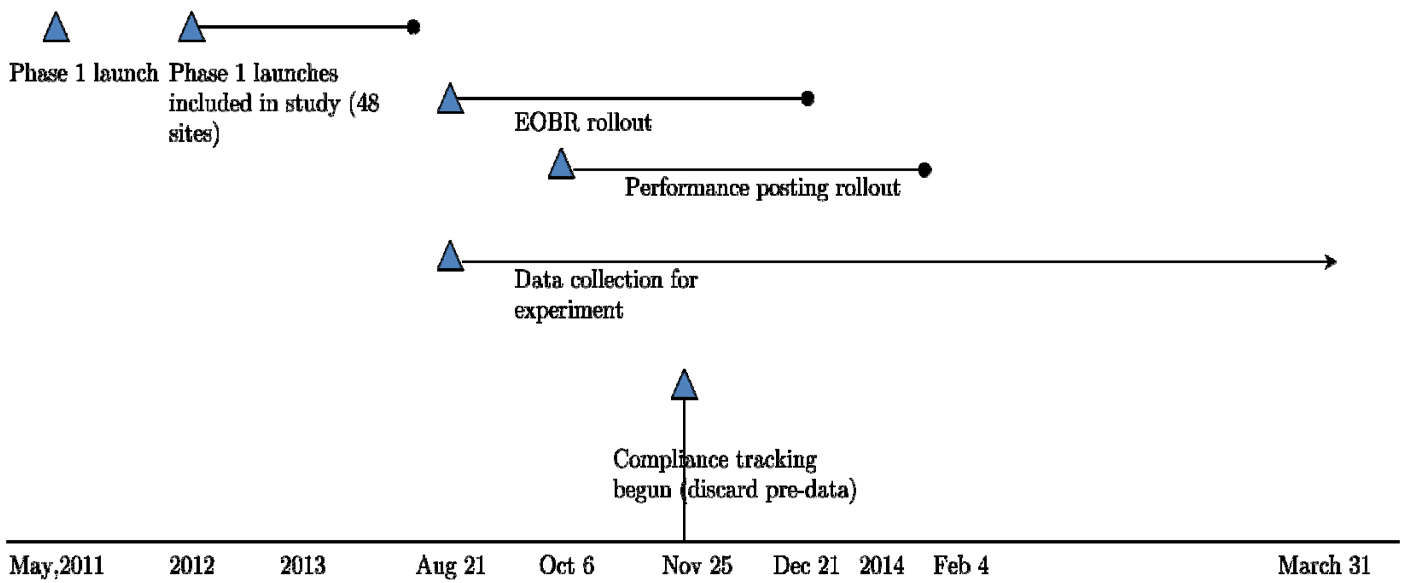
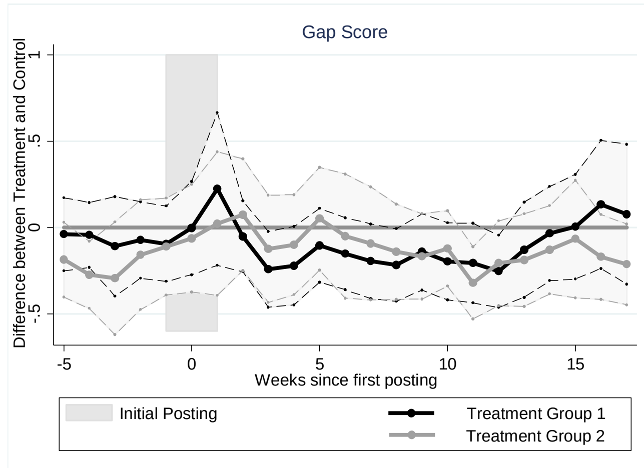
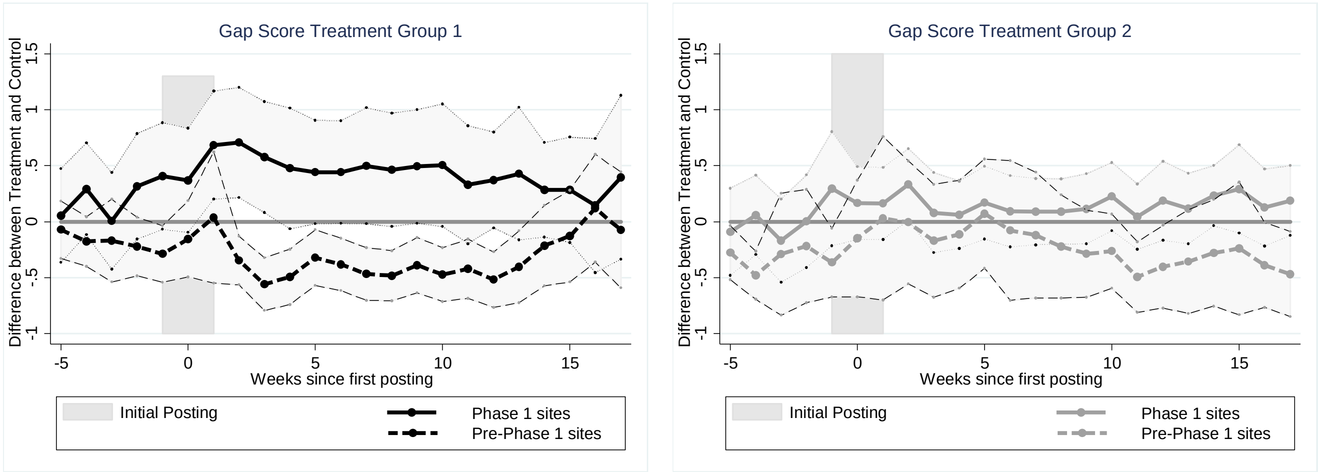


Figure 2: Impact of Rankings on Driver Performance



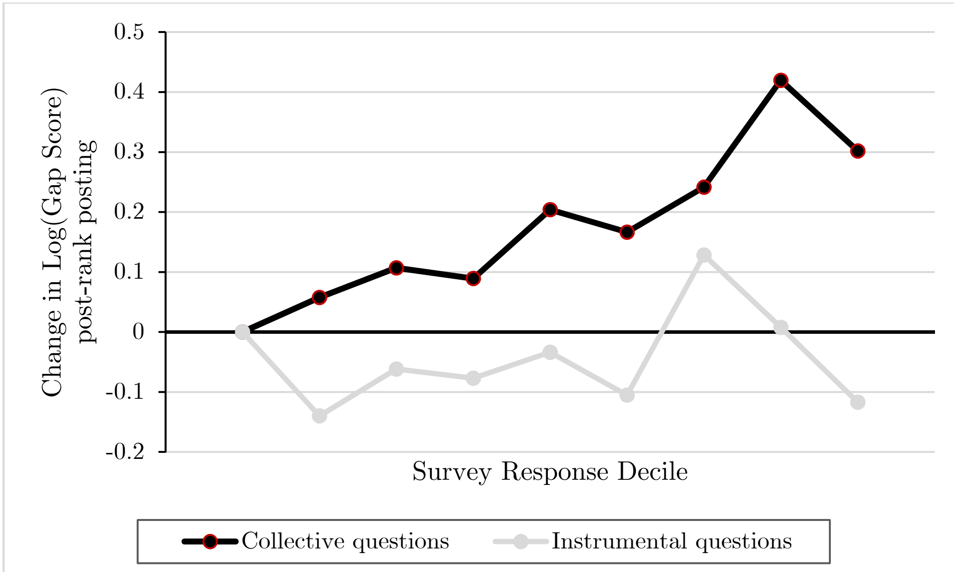
Difference between treatment and control for all sites (Phase 1 and pre-Phase 1 sites pooled together). See caption to Table 1 for definition of variables. Error bars reflect 90% confidence intervals, clustered by site. See Appendix Figure A4 for other outcome measures.

Figure 3: Impact of Rankings by Site Type



Difference between treatment and control for Phase 1 sites (left hand column) and pre-Phase 1 sites (right hand column). See caption to Table 1 for definition of variables. Error bars reflect 90% confidence intervals, clustered by site. See Appendix Figure A5 for other outcome measures.

Figure 4: Rank Posting Response by Engagement Survey Response



Difference, relative to control group, in performance outcomes between pre- and post-periods, by decile of survey response questions. Deciles are based on two indices of survey questions constructed to capture, respectively, collective identification (or team affiliation) and instrumental job satisfaction, such as satisfaction with pay and benefits.

Tables

Table 1: Balance Between Control and Treatment Groups

	Full sample					Matched sample				
	Control Mean	Treatment 1 (names) Mean	Diff p-value	Treatment 2 (IDs) Mean	Diff p-value	Control Mean	Treatment 1 (names) Mean	Diff p-value	Treatment 2 (IDs) Mean	Diff p-value
Site characteristics										
# sites	47.00	50.00	n/a	45.00	n/a	19.00	25.00	n/a	27.00	n/a
Phase 1 status	0.30	0.26	0.681	0.47	0.098	0.47	0.40	0.635	0.59	0.437
Tractors / site	25.00	25.32	0.924	23.73	0.664	27.05	31.28	0.329	25.19	0.610
Distance / trip	124.08	130.63	0.309	131.24	0.247	127.39	122.67	0.566	131.38	0.606
Eastern region	0.44	0.44	0.966	0.30	0.149	0.46	0.52	0.682	0.29	0.188
Central region	0.33	0.22	0.220	0.39	0.607	0.42	0.22	0.131	0.45	0.833
Western region	0.22	0.34	0.207	0.32	0.313	0.12	0.26	0.197	0.26	0.180
Pre-treatment driver performance										
Miles per gallon	6.76	6.88	0.247	6.82	0.558	6.62	6.91	0.073	6.79	0.236
Gap score	2.18	2.14	0.787	1.98	0.310	1.99	2.25	0.283	1.86	0.473
Shift score	90.77	90.69	0.902	91.79	0.149	91.23	90.42	0.412	92.34	0.100
Excess idle time	0.12	0.12	0.838	0.14	0.429	0.12	0.12	0.998	0.12	0.981
Fuel lost	0.34	0.35	0.722	0.31	0.185	0.34	0.35	0.653	0.31	0.221

Phase 1 status is an indicator variable that equals one if site has launched the first phase of its business transformation initiative after 2011, the initial year, and at least three months before the commencement of the experiment. *Tractors/site* is the number of tractors assigned to a given site, a measure of the size of the establishment. *Distance / trip* is the average number of miles per trip, a measure of the nature of the route (shorter routes typically denote more urban settings). *Miles per gallon* is the average actual miles per gallon for a given route. *Gap score* is the difference between potential MPG and actual MPG, a measure of driving efficiency (lower scores indicate more efficient driving). *Shift score* is the percent of shifting events during a trip that occurred within the EOBR-specified RPM range of the engine. Higher scores denote better performance, with a maximum possible score of 100. We reversed-score this metric in our analyses to conform to the other outcome variables, where higher scores denote worse performance. *Excess idle time* is a measure, in hours, of the time spent on a given route that the engine idled longer than a threshold period designated by the EOBR period. *Fuel lost* is the total calculated fuel lost in gallons / hour during a given route, based on the combination of shifting behavior, idling, driving behavior and engine calibration. Site characteristics differences calculated at site level. Pre-treatment driver performance differences calculated at driver level. The matched sample was created using a propensity-scored approach that matched Phase 1 and pre-Phase 1 sites based on *Tractors / site*, *distance / trip* and pre-treatment measures of *miles per gallon*, *Gap score* and *Shift score*.

Table 2: Balance Between Phase 1 and Pre-Phase 1 Sites

	Full sample			Matched sample		
	Pre-Phase 1 Mean	Phase 1 Mean	Diff p-value	Pre-Phase 1 Mean	Phase 1 Mean	Diff *p*-value
Site characteristics						
# sites	95	48	n/a	41	41	n/a
Tractors / site	20.35	33.25	0.000	25.95	27.51	0.581
Distance / trip	128.04	127.53	0.609	128.04	127.53	0.937
Eastern region	0.27	0.39	0.155	0.37	0.38	0.865
Central region	0.41	0.37	0.626	0.44	0.38	0.626
Western region	0.32	0.24	0.357	0.20	0.23	0.701
Control group	0.35	0.29	0.480	0.39	0.27	0.245
Treatment Group 1	0.39	0.27	0.149	0.32	0.24	0.467
Treatment Group 2	0.26	0.44	0.027	0.29	0.49	0.072
Pre-treatment driver performance						
Miles per gallon	6.90	6.72	0.039	6.76	6.71	0.602
Gap score	2.14	2.04	0.537	2.00	2.03	0.838
Shift score	90.35	91.55	0.076	91.62	91.66	0.950
Excess idle time	0.12	0.13	0.781	0.12	0.13	0.815
Fuel lost	0.34	0.33	0.473	0.32	0.33	0.753

See Table 1 caption for variable definitions.

Table 3: Effect of Rankings on All Sites

Dependent variabe:	Log(Gap Score)		Shift Score		Log(Idle Time)		Log(Fuel Lost)	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Post*Treatment Group 1	0.0016	0.0094	0.5293	0.6395*	-0.0015	-0.0008	-0.0008	0.0027
	(0.0221)	(0.0208)	(0.3593)	(0.3370)	(0.0073)	(0.0071)	(0.0076)	(0.0072)
Post*Treatment Group 2	0.0295	0.0453**	0.3932	0.5246	0.0101	0.0092	0.0094	0.0119
	(0.0263)	(0.0221)	(0.3742)	(0.3508)	(0.0073)	(0.0071)	(0.0083)	(0.0077)
Post	-0.0186	-0.0052	-0.5163	-0.3326	-0.0018	-0.0012	-0.0036	0.0024
	(0.0323)	(0.0267)	(0.4012)	(0.3677)	(0.0080)	(0.0076)	(0.0098)	(0.0096)
Treatment Group 1 (names)	-0.0081	-0.0087	-0.0704	0.0202	-0.0065	-0.0093	0.0061	-0.0022
	(0.0421)	(0.0268)	(0.5027)	(0.3668)	(0.0078)	(0.0078)	(0.0113)	(0.0101)
Treatment Group 2 (IDs)	-0.0708	-0.0134	-0.7905	-0.1254	0.0062	0.0114	-0.0131	-0.0040
	(0.0470)	(0.0372)	(0.4798)	(0.4329)	(0.0081)	(0.0074)	(0.0129)	(0.0145)
Constant	0.9478***	2.1626	8.0190***	2.8225	0.1177***	0.2988	0.2634***	0.0208
	(0.0327)	(0.1247)	(0.3352)	(1.4176)	(0.0056)	(0.0366)	(0.0080)	(0.0476)
Controls	N	Y	N	Y	N	Y	N	Y
Observations	93,913	93,913	93,913	93,913	93,913	93,913	93,913	93,913
# Drivers	4,827	4,827	4,827	4,827	4,827	4,827	4,827	4,827
# Sites (clusters)	143	143	143	143	143	143	143	143
Adjusted R-squared	0.017	0.130	0.008	0.047	0.024	0.059	0.007	0.105

For all dependent variables, lower scores represent better performance (more efficient driving). See Table 1 for variable definitions. Controls include *Tractors/site, Distance/trip, potential MPG,* and fixed effects for the Phase 1 regional manager (the manager in charge of overseeing the Phase 1 implementation), region and calendar date. Outcome variables winsorized at 1%. Sample includes thirty days before to after the scheduled posting date, excluding the 5 days before and after posting date. Standard errors clustered by site. ***, **, * indicate significance at 1%, 5% and 10%.

Table 4: Effect of Rankings on Phase 1 and Pre-Phase 1 Sites

Dependent variable:	Log(Gap Score)		Shift Score		Log(Idle Time)		Log(Fuel Lost)	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Post*Treatment Group 1*Phase 1	0.1351***	0.1364***	1.8567**	1.9631***	0.0397***	0.0354***	0.0607***	0.0549***
	(0.0406)	(0.0375)	(0.7414)	(0.6711)	(0.0125)	(0.0122)	(0.0136)	(0.0121)
Post*Treatment Group 2*Phase 1	0.0207	0.0312	0.5047	0.7213	-0.0096	-0.0098	0.0078	0.0129
	(0.0498)	(0.0451)	(0.7248)	(0.6709)	(0.0157)	(0.0146)	(0.0162)	(0.0158)
Post*Treatment Group 1	-0.0475*	-0.0400*	-0.1718	-0.0960	-0.0149*	-0.0129	-0.0224***	-0.0169**
	(0.0257)	(0.0236)	(0.4198)	(0.3900)	(0.0079)	(0.0078)	(0.0079)	(0.0081)
Post*Treatment Group 2	0.0224	0.0309	0.2686	0.2620	0.0152	0.0144	0.0066	0.0059
	(0.0370)	(0.0331)	(0.4214)	(0.4163)	(0.0128)	(0.0122)	(0.0112)	(0.0119)
Post*Phase 1	-0.0393	-0.0295	-0.9703	-0.9613*	0.0024	0.0016	-0.0122	-0.0096
	(0.0332)	(0.0298)	(0.6016)	(0.5432)	(0.0084)	(0.0082)	(0.0114)	(0.0100)
Treatment Group 1*Phase 1	0.1506*	0.0435	-0.4987	-0.7301	0.0088	-0.0141	0.0147	0.0072
	(0.0814)	(0.0645)	(0.9923)	(0.7370)	(0.0156)	(0.0148)	(0.0228)	(0.0238)
Treatment Group 2*Phase 1	0.0629	0.0897	0.8050	0.4052	0.0029	0.0235*	-0.0004	0.0313
	(0.0871)	(0.0719)	(0.9163)	(0.7771)	(0.0159)	(0.0131)	(0.0246)	(0.0278)
Post	0.0011	0.0048	-0.1242	0.0254	-0.0034	-0.0025	0.0021	0.0055
	(0.0349)	(0.0291)	(0.4080)	(0.3717)	(0.0087)	(0.0084)	(0.0110)	(0.0106)
Phase 1	-0.0914	-0.0872*	-0.0905	-0.2886	0.0011	-0.0179*	-0.0116	-0.0279
	(0.0565)	(0.0449)	(0.6810)	(0.5884)	(0.0109)	(0.0092)	(0.0154)	(0.0169)
Treatment Group 1 (names)	-0.0649	-0.0295	0.0971	0.2240	-0.0095	-0.0061	0.0005	-0.0066
	(0.0550)	(0.0340)	(0.6823)	(0.4844)	(0.0102)	(0.0096)	(0.0143)	(0.0131)
Treatment Group 2 (IDs)	-0.0908	-0.0533	-1.2382**	-0.3270	0.0044	0.0007	-0.0109	-0.0182
	(0.0640)	(0.0434)	(0.5300)	(0.5446)	(0.0108)	(0.0105)	(0.0157)	(0.0168)
Constant	0.9806***	2.1698	8.0706***	2.4119	0.1170***	0.2931	0.2672***	0.0208
	(0.0453)	(0.1292)	(0.4212)	(1.4685)	(0.0081)	(0.0371)	(0.0098)	(0.0488)
Controls	N	Y	N	Y	N	Y	N	Y
Observations	93,913	93,913	93,913	93,913	93,913	93,913	93,913	93,913
# Drivers	4,827	4,827	4,827	4,827	4,827	4,827	4,827	4,827
# Sites (clusters)	143	143	143	143	143	143	143	143
Adjusted R-squared	0.023	0.132	0.009	0.048	0.026	0.060	0.010	0.106

For all dependent variables, lower scores represent better performance (more efficient driving). See Table 1 for variable definitions and Table 3 for list of controls and description of basic specification, Errors clustered by site. ***, **, * indicate significance at 1%, 5% and 10%.

Table 5: Analysis Using Driver Fixed Effects

Dependent variable:	Log(Gap Score)		Shift Score		Log(Idle Time)		Log(Fuel Lost)	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Post*Treatment Group 1*Phase 1	0.1026***	0.1067***	1.9433***	1.9459***	0.0279**	0.0254*	0.0501***	0.0441***
	(0.0359)	(0.0355)	(0.6079)	(0.6041)	(0.0140)	(0.0132)	(0.0141)	(0.0124)
Post*Treatment Group 2*Phase 1	0.0544	0.0561	1.2449**	1.2635**	-0.0087	-0.0069	0.0169	0.0192
	(0.0396)	(0.0389)	(0.5879)	(0.5845)	(0.0157)	(0.0152)	(0.0143)	(0.0136)
Post*Treatment Group 1	-0.0424*	-0.0452**	-0.3998	-0.3974	-0.0133	-0.0109	-0.0231***	-0.0180**
	(0.0229)	(0.0221)	(0.3492)	(0.3484)	(0.0087)	(0.0084)	(0.0079)	(0.0077)
Post*Treatment Group 2	0.0112	0.0115	-0.0615	-0.0674	0.0143	0.0131	0.0042	0.0021
	(0.0304)	(0.0292)	(0.3261)	(0.3252)	(0.0127)	(0.0126)	(0.0099)	(0.0104)
Post*Phase 1	-0.0368	-0.0415	-1.2533**	-1.2798***	0.0062	0.0052	-0.0130	-0.0126
	(0.0272)	(0.0271)	(0.4947)	(0.4896)	(0.0094)	(0.0090)	(0.0105)	(0.0092)
Constant	0.9495***	1.8575***	7.081***	11.3235***	0.1260***	0.1821***	0.2564***	-0.0576***
	(0.0119)	(0.0484)	(0.1367)	(0.5761)	(0.0057)	(0.0131)	(0.0052)	(0.0195)
Controls	N	Y	N	Y	N	Y	N	Y
Observations	93,913	93,913	93,913	93,913	93,913	93,913	93,913	93,913
# Drivers	4,827	4,827	4,827	4,827	4,827	4,827	4,827	4,827
# Sites (clusters)	143	143	143	143	143	143	143	143
Adjusted R-squared	0.546	0.559	0.601	0.602	0.285	0.298	0.459	0.509

For all dependent variables, lower scores represent better performance (more efficient driving). See Table 1 for variable definitions and Table 3 for list of controls and description of basic specification. Fixed effects for calendar date and driver included. Errors clustered by site. ***, **, * indicate significance at 1%, 5% and 10%.

Table 6: Proportional Selection Bias Analysis

Panel A

	Log(Gap Score)		Shift Score		Log(Idle Time)		Log(Fuel Lost)	
$\Pi =$	1.3	2.0	1.3	2.0	1.3	2.0	1.3	2.0
δ	-11.0562	-5.5813	18.4992	5.8832	3.8645	1.5422	-5.5741	-2.8532
β^*	0.1452	0.1570	-0.4147	-0.3639	0.0182	0.0086	0.0600	0.0687
β^{tilde}	0.14032	0.14032	-0.4347	-0.4347	0.0220	0.0220	0.0564	0.0564

Panel B

	Log(Gap Score)		Shift Score		Log(Idle Time)		Log(Fuel Lost)	
$\Pi =$	1.3	2.0	1.3	2.0	1.3	2.0	1.3	2.0
δ	-18.0279	-9.6608	-3.0233	-1.2351	2.7368	1.1433	-3.3333	-1.9145
β^*	0.1684	0.1760	-1.0217	-1.3893	0.0211	0.0042	0.0717	0.0839
β^{tilde}	0.16522	0.16522	-0.862574	-0.8625736	0.0276	0.0276	0.0667	0.0667

Displays two output statistics from Oster (2014) proportional selection bias analysis. δ is the estimated coefficient of proportionality for $\beta=0$. $|\delta|>0$ occurs when omitted variable bias has at least as much effect on outcome as observed variables and are considered sufficiently unlikely to rule out omitted variable driving the observed treatment effect (Oster 2014). β^* is the estimated treatment effect given $\delta=1$, or equal selection on observables and unobservables. Panel A uses a diffs-in-diffs specification where the treatment effect is *Phase 1*Treatment Group 1* on a sample that includes all post-period observations for the control and the Treatment Group 1 (Named Postings). Panel B uses an OLS model where the treatment is *Phase 1* and the sample is all post-period observations in Treatment Group 1.

Table 7: Matched Analysis

Dependent variable:	Log(Gap Score)		Shift Score		Log(Idle Time)		Log(Fuel Lost)	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Post*Treatment Group 1*Phase 1	0.1356***	0.1425***	-0.2841	0.0365	0.0480***	0.0482***	0.0549***	0.0561***
	(0.0504)	(0.0433)	(0.7637)	(0.7349)	(0.0167)	(0.0152)	(0.0155)	(0.0144)
Post*Treatment Group 2*Phase 1	-0.0131	0.0126	-0.5912	-0.3145	-0.0276	-0.0212	0.0006	0.0048
	(0.0606)	(0.0509)	(0.6610)	(0.6081)	(0.0184)	(0.0167)	(0.0183)	(0.0189)
Post*Treatment Group 1	-0.0672*	-0.0563*	0.8533	0.7136	-0.0239**	-0.0239**	-0.0257**	-0.0220*
	(0.0359)	(0.0332)	(0.5163)	(0.5417)	(0.0109)	(0.0108)	(0.0115)	(0.0112)
Post*Treatment Group 2	0.0332	0.0341	0.5088	0.4096	0.0286*	0.0240	0.0069	0.0081
	(0.0491)	(0.0430)	(0.4415)	(0.4429)	(0.0157)	(0.0151)	(0.0152)	(0.0156)
Post*Phase 1	-0.0451	-0.0404	0.2467	0.1883	-0.0021	-0.0060	-0.0119	-0.0120
	(0.0394)	(0.0318)	(0.4928)	(0.4763)	(0.0107)	(0.0105)	(0.0119)	(0.0119)
Treatment Group 1*Phase 1	0.0837	0.0470	-0.1255	0.8650	0.0102	0.0055	-0.0067	0.0185
	(0.1024)	(0.0887)	(1.0052)	(0.9060)	(0.0192)	(0.0219)	(0.0260)	(0.0317)
Treatment Group 2*Phase 1	-0.0625	0.1101	0.4843	1.6397*	-0.0009	0.0361**	-0.0323	0.0456
	(0.1102)	(0.0873)	(0.9885)	(0.9518)	(0.0172)	(0.0175)	(0.0299)	(0.0323)
Post	0.0090	0.0045	-0.9246*	-0.4917	0.0005	0.0012	0.0056	0.0041
	(0.0466)	(0.0392)	(0.4957)	(0.4563)	(0.0122)	(0.0121)	(0.0145)	(0.0143)
Phase 1	-0.0016	-0.0448	0.2930	-0.8180	-0.0024	-0.0201	0.0158	-0.0219
	(0.0767)	(0.0655)	(0.7303)	(0.7741)	(0.0115)	(0.0131)	(0.0185)	(0.0247)
Treatment Group 1 (names)	-0.0272	0.0135	-0.0684	-0.2473	-0.0023	-0.0033	0.0027	0.0095
	(0.0660)	(0.0624)	(0.6681)	(0.6186)	(0.0141)	(0.0181)	(0.0162)	(0.0238)
Treatment Group 2 (IDs)	0.0459	-0.0908	-0.6934	-1.2325	0.0113	-0.0098	0.0109	-0.0319
	(0.0783)	(0.0678)	(0.5594)	(0.8952)	(0.0130)	(0.0145)	(0.0199)	(0.0257)
Constant	0.9032***	2.4719***	7.3636***	13.9057***	0.1170***	0.3254***	0.2606***	0.1451**
	(0.0595)	(0.1654)	(0.5236)	(1.3605)	(0.0127)	(0.0404)	(0.0166)	(0.0589)
Controls	N	Y	N	Y	N	Y	N	Y
Observations	60,002	60,002	60,002	60,002	60,002	60,002	60,002	60,002
# Drivers	3,065	3,065	3,065	3,065	3,065	3,065	3,065	3,065
# Sites (clusters)	82	82	82	82	82	82	82	82
Adjusted R-squared	0.021	0.130	0.012	0.052	0.031	0.066	0.010	0.104

For all dependent variables, lower scores represent better performance (more efficient driving). See Table 1 for variable definitions and Table 3 for list of controls and description of basic specification. See Tables 1 and 2 for description of the matched sample. Errors clustered by site. ***, **, * indicate significance at 1%, 5% and 10%.

Table 8: Matched Analysis with Driver Fixed Effects

Dependent variable:	Log(Gap Score)		Shift Score		Log(Idle Time)		Log(Fuel Lost)	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Post*Treatment Group 1*Phase 1	0.1088***	0.1092***	0.1514	0.1495	0.0360**	0.0350**	0.0453***	0.0435***
	(0.0392)	(0.0380)	(0.6861)	(0.6841)	(0.0172)	(0.0161)	(0.0152)	(0.0135)
Post*Treatment Group 2*Phase 1	0.0118	0.0214	0.0348	0.0812	-0.0267	-0.0234	0.0031	0.0043
	(0.0436)	(0.0426)	(0.5283)	(0.5263)	(0.0179)	(0.0175)	(0.0163)	(0.0158)
Post*Treatment Group 1	-0.0525*	-0.0514*	0.4054	0.4153	-0.0215*	-0.0196*	-0.0231**	-0.0204*
	(0.0308)	(0.0303)	(0.4953)	(0.4945)	(0.0114)	(0.0110)	(0.0114)	(0.0107)
Post*Treatment Group 2	0.0317	0.0274	0.1981	0.1755	0.0280*	0.0259*	0.0096	0.0080
	(0.0397)	(0.0388)	(0.3252)	(0.3275)	(0.0152)	(0.0152)	(0.0137)	(0.0141)
Post*Phase 1	-0.0299	-0.0355	-0.0595	-0.0869	0.0026	0.0006	-0.0078	-0.0086
	(0.0279)	(0.0268)	(0.4627)	(0.4589)	(0.0112)	(0.0109)	(0.0110)	(0.0102)
Constant	0.9295***	1.3188***	7.5875***	7.0628***	0.1297***	0.2249***	0.2558***	-0.2134***
	(0.0161)	(0.0701)	(0.1689)	(0.7805)	(0.0073)	(0.0191)	(0.0072)	(0.0254)
Controls	N	Y	N	Y	N	Y	N	Y
Driver Fixed Effects	Y	Y	Y	Y	Y	Y	Y	Y
Observations	60,002	60,002	60,002	60,002	60,002	60,002	60,002	60,002
# Drivers	3,065	3,065	3,065	3,065	3,065	3,065	3,065	3,065
# Sites (clusters)	82	82	82	82	82	82	82	82
Adjusted R-squared	0.559	0.572	0.620	0.621	0.289	0.303	0.470	0.521

For all dependent variables, lower scores represent better performance (more efficient driving). This table contains the same specification as Table 5, but on the matched sample. See Table 1 for variable definitions and Table 3 for list of controls and description of basic specification. See Tables 1 and 2 for description of the matched sample. Errors clustered by site. ***, **, * indicate significance at 1%, 5% and 10%.

Table 9: Effect on Variance

Coefficient of variation:	Log(Gap Score)		Shift Score		Log(Idle Time)		Log(Fuel Lost)	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Post*Treatment Group 1*Phase 1	-0.0427*	-0.0427**	-0.0657**	-0.0651**	-0.2466***	-0.2240***	-0.0184	-0.0175
	(0.0239)	(0.0202)	(0.0332)	(0.0293)	(0.0787)	(0.0677)	(0.0354)	(0.0280)
Post*Treatment Group 2*Phase 1	0.0042	-0.0036	-0.0333	-0.0461	-0.0484	-0.0509	0.0265	0.0078
	(0.0241)	(0.0213)	(0.0338)	(0.0292)	(0.0776)	(0.0641)	(0.0339)	(0.0288)
Post*Treatment Group 1	0.0389***	0.0288**	0.0028	-0.0112	0.1335***	0.0890**	0.0311*	0.0206
	(0.0129)	(0.0117)	(0.0189)	(0.0169)	(0.0421)	(0.0358)	(0.0177)	(0.0158)
Post*Treatment Group 2	-0.0169	-0.0118	-0.0226	-0.0174	-0.0171	0.0113	-0.0420*	-0.0297
	(0.0166)	(0.0144)	(0.0216)	(0.0182)	(0.0516)	(0.0398)	(0.0231)	(0.0194)
Post*Phase	-0.0059	-0.0023	0.0156	0.0204	0.0867	0.1058**	-0.0037	0.0022
	(0.0155)	(0.0142)	(0.0230)	(0.0194)	(0.0543)	(0.0467)	(0.0230)	(0.0196)
Treatment Group 1*Phase 1	-0.0034	-0.0086	0.0986***	0.0610***	-0.0029	-0.0399	0.0038	-0.0281
	(0.0171)	(0.0154)	(0.0236)	(0.0219)	(0.0591)	(0.0530)	(0.0249)	(0.0209)
Treatment Group 2*Phase 1	-0.0573***	-0.0175	0.0330	0.1100***	-0.1694***	-0.1905***	-0.0463*	0.0135
	(0.0173)	(0.0174)	(0.0239)	(0.0233)	(0.0581)	(0.0536)	(0.0244)	(0.0235)
Post	-0.0253***	-0.0217**	0.0262**	-0.0014	-0.2051***	-0.0910***	-0.0180	-0.0139
	(0.0087)	(0.0092)	(0.0134)	(0.0136)	(0.0289)	(0.0284)	(0.0120)	(0.0124)
Phase	0.0493***	0.0233*	0.0004	-0.0471***	0.1039**	0.0898**	0.0447***	0.0197
	(0.0111)	(0.0120)	(0.0161)	(0.0160)	(0.0405)	(0.0415)	(0.0162)	(0.0165)
Treatment Group 1 (names)	0.0149	0.0114	0.0149	0.0276**	-0.0476	-0.0075	-0.0044	0.0121
	(0.0096)	(0.0090)	(0.0136)	(0.0130)	(0.0317)	(0.0282)	(0.0130)	(0.0121)
Treatment Group 2 (IDs)	0.0729***	0.0321***	0.0258*	0.0069	0.0732*	0.0175	0.0463***	0.0005
	(0.0122)	(0.0122)	(0.0156)	(0.0154)	(0.0393)	(0.0335)	(0.0172)	(0.0163)
Constant	0.5712***	0.3294***	0.8069***	0.8392***	1.6574***	0.1995	0.7965***	1.1555***
	(0.0066)	(0.0720)	(0.0095)	(0.0971)	(0.0224)	(0.1910)	(0.0086)	(0.0981)
Controls	N	Y	N	Y	N	Y	N	Y
Observations	5,343	5,343	5,343	5,343	5,343	5,343	5,343	5,343
# Sites (clusters)	143	143	143	143	143	143	143	143
Adjusted R-squared	0.024	0.240	0.022	0.192	0.029	0.284	0.007	0.280

Dependent variable is the coefficient of variation of the outcome measures. Observations represent site-date combinations. See Table 3 for list of controls. Controls that are measured at a driver-date level (*Tractors/site, Distance/trip, potential MPG*) have been averaged to the site level for this specification. Errors clustered by site. ***, **, * indicate significance at 1%, 5% and 10%.

Table 10: Effect of Ranking and Engagement on Driver Performance

Dependent variable:	Log(Gap Score)			
Category:	Collective Index		Instrumental Index	
	(1)	(2)	(3)	(4)
Post*Treatment Group 1*[Category]	0.1101**	0.1146**	0.0669	0.0751
	(0.0425)	(0.0543)	(0.0435)	(0.0572)
Post*Treatment Group 2*[Category]	0.0300	0.0046	-0.0327	-0.1055
	(0.0620)	(0.0811)	(0.0762)	(0.1226)
Post*Treatment Group 1	-0.3991***	-0.4352**	-0.2732*	-0.3214*
	(0.1279)	(0.1567)	(0.1365)	(0.1745)
Post*Treatment Group 2	-0.0506	-0.0556	0.1582	0.3164
	(0.2226)	(0.3156)	(0.2731)	(0.4899)
Post*[Category]	-0.0587*	-0.0553	-0.0293	-0.0322
	(0.0337)	(0.0470)	(0.0321)	(0.0461)
Treatment Group 1*[Category]	0.1184*	0.1353*	0.0917	0.1187
	(0.0606)	(0.0666)	(0.0728)	(0.0848)
Treatment Group 2*[Category]	-0.0567	0.0216	-0.0505	0.0079
	(0.0685)	(0.0819)	(0.0677)	(0.0945)
Post	0.1881	0.3620*	0.1558	0.4125
	(0.1626)	(0.1979)	(0.1859)	(0.2816)
[Category]	-0.0341	-0.0488	-0.0316	-0.0352
	(0.0405)	(0.0413)	(0.0488)	(0.0542)
Treatment Group 1 (Names)	-0.3244	-0.3533	-0.2334	-0.2874
	(0.2155)	(0.2324)	(0.2680)	(0.3040)
Treatment Group 2 (IDs)	0.2840	0.0538	0.2735	0.1122
	(0.2136)	(0.2602)	(0.2175)	(0.3162)
Constant	2.1588***	2.1343***	1.9907***	1.8066***
	(0.3503)	(0.3496)	(0.3570)	(0.3598)
Controls	Y	Y	Y	Y
Sample	Full	Matched	Full	Matched
Observations	35,187	26,065	35,187	26,065
# Drivers	491	368	491	368
# Sites (clusters)	43	26	43	26
Adjusted R-squared	0.107	0.120	0.098	0.116

Lower scores represent better performance (more efficient driving). This table reproduces the analysis of Table 3, but substituting the indices constructed from the engagement survey for Phase 1 status. Controls include both all available demographic controls from the survey (*race*, *age* and *tenure at company*), calendar date fixed effects, and the site and driver route controls from Table 3. Errors clustered by site. ***, **, * indicate significance at 1%, 5% and 10%.

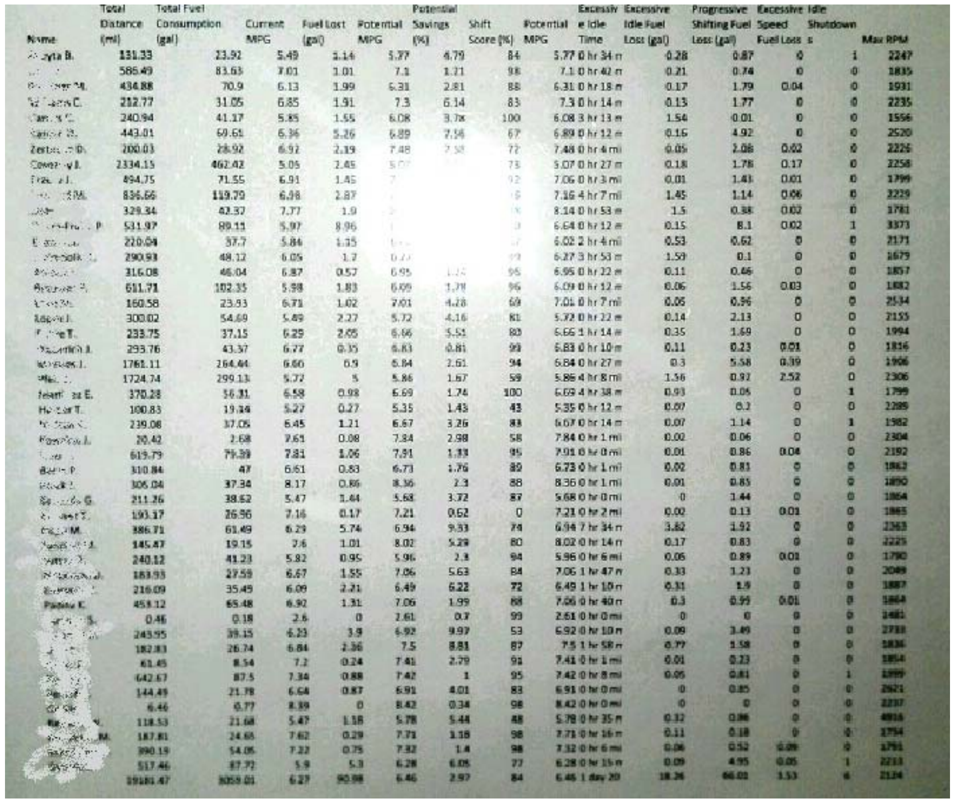
Table 11: Effect of Ranking and Engagement on Driver Performance, Driver Fixed Effects

Depedent variable:	Log(Gap Score)			
Category:	Collective Index		Instrumental Index	
	(1)	(2)	(3)	(4)
Post*Treatment Group 1*[Category]	0.0988***	0.1033**	0.0548	0.0398
	(0.0348)	(0.0461)	(0.0359)	(0.0448)
Post*Treatment Group 2*[Category]	0.0227	0.0162	-0.0333	-0.0931
	(0.0399)	(0.0543)	(0.0591)	(0.0957)
Post*Treatment Group 1	-0.3895***	-0.4021***	-0.2587**	-0.2049
	(0.1010)	(0.1342)	(0.1052)	(0.1305)
Post*Treatment Group 2	-0.0462	-0.0620	0.1339	0.3052
	(0.1283)	(0.1841)	(0.1956)	(0.3503)
Post*[Category]	-0.0534*	-0.0571	-0.0290	-0.0212
	(0.0275)	(0.0401)	(0.0299)	(0.0400)
Constant	1.7088***	1.6556***	1.6518***	1.8221***
	(0.1545)	(0.1670)	(0.2027)	(0.1795)
Controls	Y	Y	Y	Y
Sample	Full	Matched	Full	Matched
Observations	35,187	26,065	35,187	26,065
# Drivers	491	368	491	368
# Sites (clusters)	43	26	43	26
Adjusted R-squared	0.524	0.525	0.530	0.534

Lower scores represent better performance (more efficient driving). This table reproduces the analysis of Table 10, but including driver effects. Errors clustered by site. ***, **, * indicate significance at 1%, 5% and 10%.

For Online Publication: Appendix Figures and Tables

Figure A1: Sample Rank Posting



Name	Total Distance (mi)	Total Fuel Consumption (gal)	Current MPG	Fuel Lost (gal)	Potential MPG	Potential Savings (%)	Shift Score (%)	Potential MPG	Excessive Idle Time	Excessive Idle Fuel Loss (gal)	Progressive Shifting Fuel Loss (gal)	Excessive Speed Fuel Loss	Idle Shutdowns	Max RPM
[illegible]	131.33	23.92	5.49	1.14	5.77	4.79	84	5.77	0 hr 34 m	0.28	0.87	0	1	2247
[illegible]	585.49	83.63	7.01	1.01	7.1	1.21	98	7.1	0 hr 42 m	0.21	0.74	0	0	1835
[illegible]	434.88	70.9	6.13	1.99	6.31	2.81	88	6.31	0 hr 18 m	0.17	1.79	0.04	0	1931
[illegible]	212.77	31.05	6.85	1.91	7.3	6.14	83	7.3	0 hr 14 m	0.13	1.77	0	0	2235
[illegible]	240.94	41.17	5.85	1.55	6.08	3.78	100	6.08	3 hr 13 m	1.54	0.01	0	0	1556
[illegible]	443.01	69.61	6.36	5.26	6.89	7.56	67	6.89	0 hr 12 m	0.16	4.92	0	0	2520
[illegible]	200.03	28.92	6.92	2.19	7.48	7.58	72	7.48	0 hr 4 mi	0.05	2.06	0.02	0	2226
[illegible]	2334.15	462.42	5.05	2.45	[illegible]	[illegible]	73	5.07	0 hr 27 m	0.18	1.76	0.17	0	2258
[illegible]	494.75	71.55	6.91	1.45	[illegible]	[illegible]	92	7.06	0 hr 3 mi	0.01	1.41	0.01	0	1799
[illegible]	836.66	119.79	6.98	2.87	[illegible]	[illegible]	[illegible]	7.16	4 hr 7 mi	1.45	1.14	0.06	0	2229
[illegible]	329.34	42.37	7.77	1.9	[illegible]	[illegible]	[illegible]	8.14	0 hr 53 m	1.5	0.38	0.02	0	1781
[illegible]	531.97	89.11	5.97	8.96	[illegible]	[illegible]	[illegible]	6.64	0 hr 12 m	0.15	8.1	0.02	1	3373
[illegible]	220.04	37.7	5.84	1.15	[illegible]	[illegible]	[illegible]	6.02	2 hr 4 mi	0.53	0.62	0	0	2171
[illegible]	290.93	48.12	6.05	1.7	[illegible]	[illegible]	[illegible]	6.27	3 hr 53 m	1.59	0.1	0	0	1679
[illegible]	316.08	46.04	6.87	0.57	6.95	[illegible]	95	6.95	0 hr 22 m	0.11	0.46	0	0	1857
[illegible]	611.71	102.35	5.98	1.83	6.09	1.78	96	6.09	0 hr 12 m	0.06	1.56	0.03	0	1882
[illegible]	160.58	23.93	6.71	1.02	7.01	4.28	69	7.01	0 hr 7 mi	0.05	0.96	0	0	2534
[illegible]	300.02	54.69	5.49	2.27	5.72	4.16	81	5.72	0 hr 22 m	0.14	2.13	0	0	2155
[illegible]	233.75	37.15	6.29	2.65	6.66	5.51	80	6.66	1 hr 14 m	0.35	1.69	0	0	1994
[illegible]	293.76	43.37	6.77	0.35	6.83	0.81	99	6.83	0 hr 10 m	0.11	0.23	0.01	0	1816
[illegible]	1761.11	264.44	6.66	0.9	6.84	2.61	94	6.84	0 hr 27 m	0.3	5.58	0.39	0	1906
[illegible]	1724.74	299.13	5.72	5	5.86	1.67	59	5.86	4 hr 8 mi	1.56	0.97	2.52	0	2306
[illegible]	370.28	56.31	6.58	0.98	6.69	1.74	100	6.69	4 hr 38 m	0.93	0.05	0	1	1799
[illegible]	100.83	19.14	5.27	0.27	5.35	1.43	43	5.35	0 hr 12 m	0.07	0.2	0	0	2289
[illegible]	239.08	37.05	6.45	1.21	6.67	3.26	83	6.67	0 hr 14 m	0.07	1.14	0	1	1982
[illegible]	20.42	2.68	7.61	0.08	7.84	2.98	58	7.84	0 hr 1 mi	0.02	0.06	0	0	2304
[illegible]	619.79	79.39	7.81	1.06	7.91	1.33	95	7.91	0 hr 0 mi	0.01	0.86	0.04	0	2192
[illegible]	310.84	47	6.61	0.83	6.73	1.76	89	6.73	0 hr 1 mi	0.02	0.81	0	0	1862
[illegible]	306.04	37.34	8.17	0.86	8.36	2.3	88	8.36	0 hr 1 mi	0.01	0.85	0	0	1890
[illegible]	211.26	38.62	5.47	1.44	5.68	3.72	87	5.68	0 hr 0 mi	0	1.44	0	0	1864
[illegible]	193.17	26.96	7.16	0.17	7.21	0.62	0	7.21	0 hr 2 mi	0.02	0.13	0.01	0	1865
[illegible]	386.71	61.49	6.29	5.74	6.94	9.33	74	6.94	7 hr 34 m	3.82	1.92	0	0	2363
[illegible]	145.47	19.15	7.6	1.01	8.02	5.29	80	8.02	0 hr 14 m	0.17	0.83	0	0	2225
[illegible]	240.12	41.23	5.82	0.95	5.96	2.3	94	5.96	0 hr 6 mi	0.06	0.89	0.01	0	1790
[illegible]	183.93	27.59	6.67	1.55	7.06	5.63	84	7.06	1 hr 47 m	0.33	1.21	0	0	2049
[illegible]	216.09	35.49	6.09	2.21	6.49	6.22	72	6.49	1 hr 10 m	0.31	1.9	0	0	1887
[illegible]	453.12	65.48	6.92	1.31	7.06	1.99	88	7.06	0 hr 40 m	0.3	0.99	0.01	0	1864
[illegible]	0.46	0.18	2.6	0	2.61	0.7	99	2.61	0 hr 0 mi	0	0	0	0	1481
[illegible]	243.95	39.15	6.23	3.9	6.92	9.97	53	6.92	0 hr 10 m	0.09	3.49	0	0	2733
[illegible]	182.83	26.74	6.84	2.36	7.5	8.81	87	7.5	1 hr 58 m	0.77	1.58	0	0	1836
[illegible]	61.45	8.54	7.2	0.24	7.41	2.79	91	7.41	0 hr 1 mi	0.01	0.23	0	0	1854
[illegible]	642.67	87.5	7.34	0.88	7.42	1	95	7.42	0 hr 8 mi	0.05	0.81	0	1	1999
[illegible]	144.49	21.78	6.64	0.87	6.91	4.01	83	6.91	0 hr 0 mi	0	0.85	0	0	2621
[illegible]	6.46	0.77	8.39	0	8.42	0.34	98	8.42	0 hr 0 mi	0	0	0	0	2237
[illegible]	118.53	21.68	5.47	1.18	5.78	5.44	48	5.78	0 hr 35 m	0.32	0.86	0	0	4916
[illegible]	187.81	24.65	7.62	0.29	7.71	1.18	98	7.71	0 hr 16 m	0.11	0.18	0	0	1754
[illegible]	390.19	54.05	7.22	0.75	7.32	1.4	98	7.32	0 hr 6 mi	0.06	0.52	0.09	0	1791
[illegible]	517.46	87.72	5.9	5.3	6.28	6.05	77	6.28	0 hr 15 m	0.09	4.95	0.05	1	2213
	19181.47	3059.01	6.27	90.98	6.46	2.97	84	6.46	1 day 20	18.36	66.01	3.53	6	2124

This photograph shows the typical performance posting in a Treastment Group 1 site. All metrics are calculated as weekly averages since the last posting. From left to right, the columns are: driver name, average route distance, total fuel consumption, average MPG, Total fuel lost in gallons, Potential MPG, Gap score, Shift score, Potential MPG (repeated), Excess idle time, Fuel lost from excess idle time, Fuel lost from shifting, Fuel lost from excessive speeding, Idle shutdown, Max engine RPM. These last five measures are minor metrics that are captured by the EOBR system but not the subject of significant management attention and not supplied to the researchers.

Figure A2: Phase 1 Evaluation Criteria

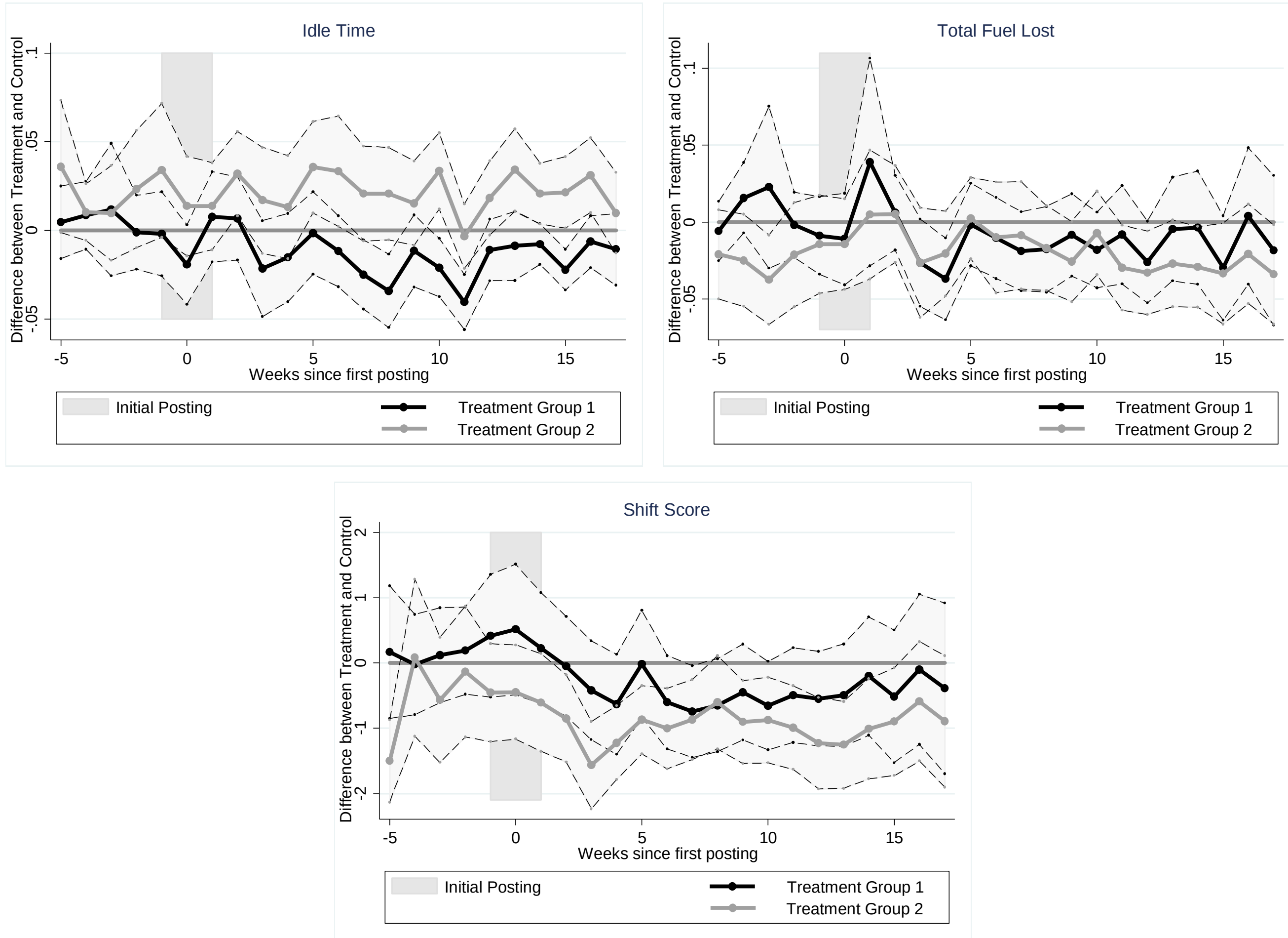
Safety	Employee s have a formal avenue to openly voice, share, and regularly address safety concerns at the facility
	Safety concerns are addressed in a timely manner by a cross-functional, integrated team of employees, supervision, and management.
Safety and leadership	What level of leader is involved in the safety journey?
	What organizational levels originated, supported, and have advocated the lean implementation initiative in the facility?
Power distance	Management availability to team members. Do employees feel that management is approachable?
	What percentage of the day do Supervisors spend on the Dock, during normal working hours?
	What percentage of the day do Managers spend on the Dock, during normal working hours?
Employee recognition	Individuals who meet, exceed, or achieve objectives are recognized on a regular basis through an employee recognition program?
	Groups who meet, exceed, or achieve objectives are recognized on a regular basis through a group recognition program?
Management style	Feedback and concerns are encouraged and included before making changes and taking actions.
	Employees, Supervisors, and Managers are encouraged/empowered to try improvement ideas, using innovation and creativity to enrich job responsibilities.
	The organizational level involved in determining and leading facility, function, and CIR Goals.
Teamwork and empowerment	Daily work activities are organized into team functions.
	SME s are utilized as initial point of contact for problem-solving, resolution, and employee directing activities.
	Problem-Solving activities are organized into team functions.
	Employees are empowered, utilized, participate, initiate, and lead problem-solving activities autonomously, without significant management involvement.
Communication	There is an avenue for workers to openly share common concerns, issues, and problems regularly with other employees, supervisors, and management.
	Employee concerns and questions are addressed in a timely manner.
	Are there daily meetings with employees and supervision/management where the daily plans, performance, etc. are shared?

These criteria are taken from a formal assessment tool used by managers to score how successful the business transformation rollout has been at any given site. Sites are assessed formally in a two-day process at least once per year to certify their progression in adopting the new culture and practices.

Figure A3: Samples from Interviews on Phase 1 of the Business Transformation Initiative

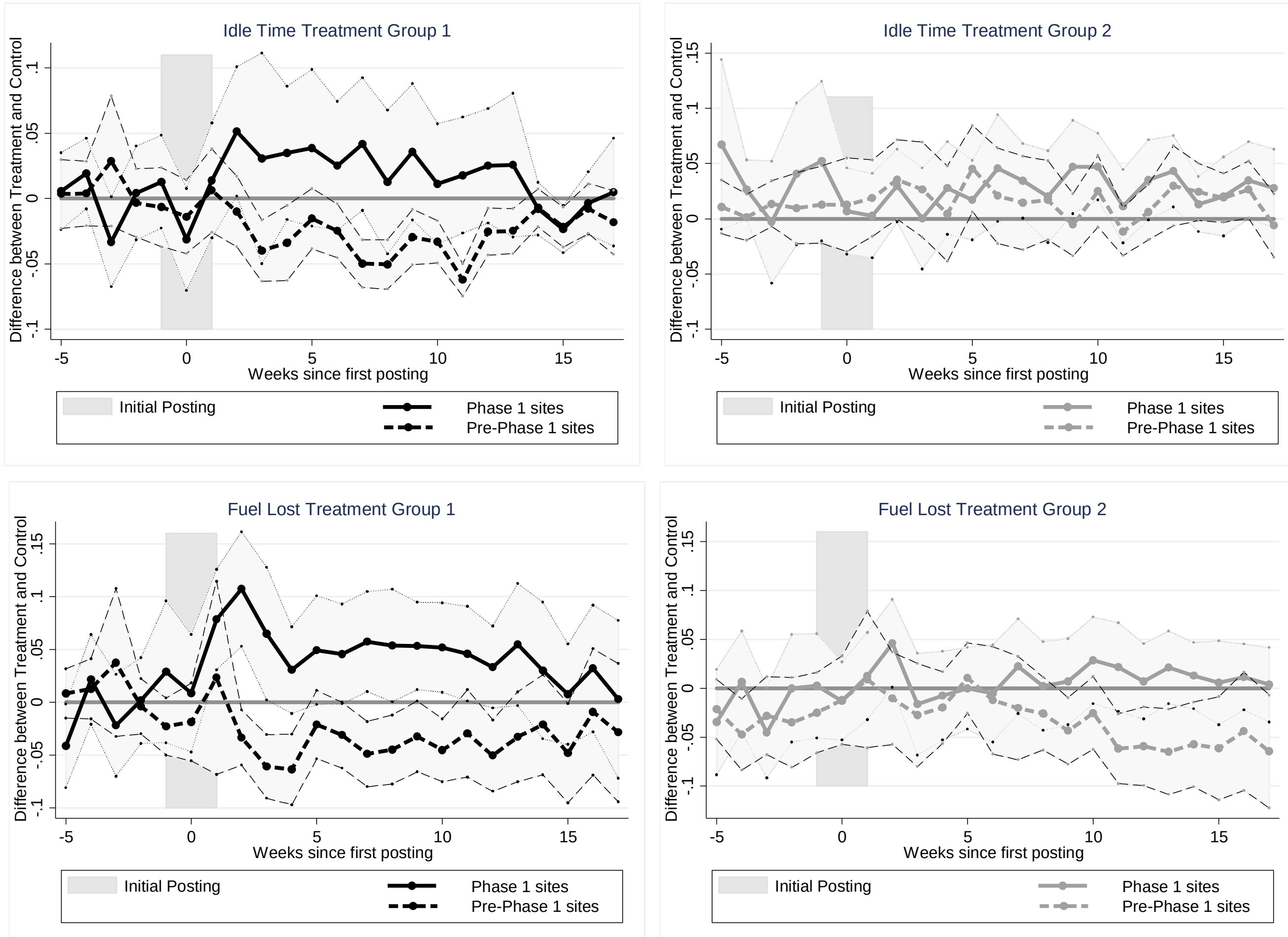
Supervisor on how initiative has affected his management style	"These guys will do anything for me, and they'll do absolutely nothing for other people. And I learned a lot of that from lean because lean has made me softer, it really has. I used to be hard as rock and now I feel like I'm a spongefiI still have that same pride but it's – my interaction with people is so much different, it's so much different. You're not treating them in a negative way or a negative manner and that's – I was hard as a rock in my numbers produced fiand if somebody didn't want to get on board with me on my team in all likelihood it probably wasn't going to be a very good day for that person. Now, it's with everybody being involved instead of just me running the show, it's totally different. Yes, are my numbers as good? Probably not, but you know what I'll take that. I firmly believe I'm a better supervisor today than what I was 6 months back."
Supervisor #2 on how initiative has motivated drivers	"Since lean was introduced it was sort of like the door opening up. [Manager said] give it a chance, look at it and see what it can do. And I tell you it can produce productivity out of people that you thought would never produce. All it takes is a little bit of respect, little bit of understanding, show these guys that they're part of the operation."
Driver #1 on how initiative has created community	"These guys now they get together, we got great relationships outside of the work environment. We've been to some of their homes. We do the activities outside of work. Even though I got Friday nights about once every month I sneak on down to Fridays and I buy them all the drink. It's just made us such a cohesive team it's incredible."
Driver #2 on how initiative has increased teamwork	"I guess we haven't really been able to do too much yet – but I think the meetings and stuff have actually helped just getting people working together. So in the lean team, I think there's actually a good amount of camaraderie going on. So I think that's actually been good. Now some people I didn't really get along and stuff are working together."

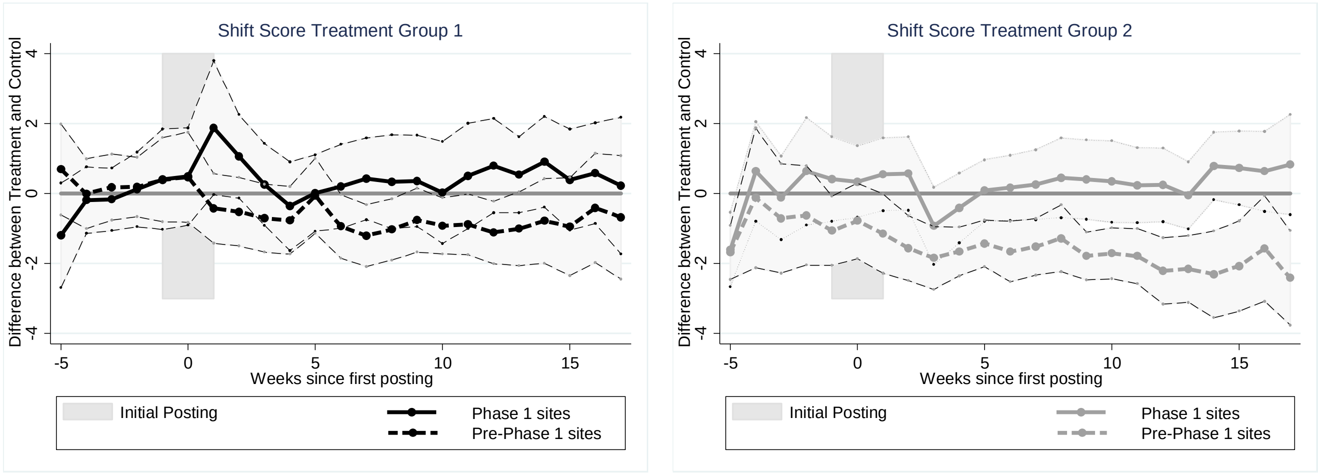
Figure A4: Impact of Rankings on Driver Performance



Difference between treatment and control for all sites (Phase 1 and pre-Phase 1 sites pooled together). See caption to Table 1 for definition of variables. Error bars reflect 90% confidence intervals, clustered by site.

Figure A5: Impact of Rankings by Site Type





Difference between treatment and control for Phase 1 sites (left hand column) and pre-Phase 1 sites (right hand column). See caption to Table 1 for definition of variables. Error bars reflect 90% confidence intervals, clustered by site.

Table A1: Sample Construction

Sample construction	Driver-days	Sites
Total driver-days	1,137,192	XXX*
- less early Phase 1 sites	(173,461)	(25)
- less late Q3/Q4 2013 Phase 1 sites	(130,679)	(XX)*
- less pre-11/25 rank posting dates	(416,593)	(72)
- less line haul routes	(76,989)	0
- less uncalibrated data	(8,781)	0
Sample	330,689	143
Sample within 5-30 window	93,913	143

*Total number of sites and sites removed has been masked for confidentiality purposes.

Table A2: Descriptive Statistics

	All Sites				Sample				Matched sample			
	Mean	Std Dev	Min	Max	Mean	Std Dev	Min	Max	Mean	Std Dev	Min	Max
Site characteristics												
# sites	XXX*	n/a	n/a	n/a	143	n/a	n/a	n/a	82	n/a	n/a	n/a
Phase 1 initiated	0.34	0.47	0	1	0.34	0.47	0	1	0.50	0.50	0	1
Tractors / site	30.86	23.52	5	151	24.58	15.36	5	87	26.73	12.70	8	61
Distance / trip	192.03	54.25	43.57	369.57	129.22	30.14	63.81	199.82	128.19	28.90	63.81	193.97
Pre-treatment driver performance												
Miles per gallon	6.74	0.57	5.25	12.63	6.81	0.51	5.71	8.63	6.70	0.47	5.71	7.74
Gap score	2.05	0.85	0.51	6.77	2.12	1.05	0.67	8.05	2.01	0.84	0.79	4.99
Shift score	91.28	3.28	75.38	97.35	91.33	5.83	41	97.82	92.46	3.27	83.94	97.72
Excess idle time	0.11	0.08	0.01	0.74	0.14	0.09	0.01	0.74	0.14	0.09	0.04	0.74
Fuel lost	0.43	0.17	0.14	1.42	0.34	0.11	0.09	1.02	0.32	0.10	0.14	0.68

* Total number of sites masked for confidentiality purposes. Left hand columns display descriptive statistics across all sites within the company. The center columns provide statistics for the sites in our experimental sample and the right hand columns provide statistics for sites iwhtin the propensity-score matched subsample.

Table A3: Table 3 Replicated on Matched Sample

Dependent variabe:	Log(Gap Score)		Shift Score		Log(Idle Time)		Log(Fuel Lost)	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Post*Treatment Group 1	-0.0028	0.0117	0.5636	0.6420*	-0.0003	0.0001	-0.0008	0.0040
	(0.0286)	(0.0272)	(0.3859)	(0.3768)	(0.0109)	(0.0101)	(0.0099)	(0.0098)
Post*Treatment Group 2	0.0258	0.0446*	-0.0212	0.1397	0.0160*	0.0152*	0.0075	0.0117
	(0.0311)	(0.0258)	(0.3047)	(0.2833)	(0.0094)	(0.0091)	(0.0096)	(0.0092)
Post	0.0324	0.0173	0.3668	0.1789	0.0239***	0.0193***	0.0123*	0.0060
	(0.0206)	(0.0181)	(0.2346)	(0.2348)	(0.0063)	(0.0062)	(0.0062)	(0.0066)
Treatment Group 1 (names)	0.0001	0.0206	-0.1697	0.0019	-0.0007	-0.0064	-0.0024	0.0138
	(0.0558)	(0.0441)	(0.5326)	(0.4811)	(0.0128)	(0.0124)	(0.0137)	(0.0163)
Treatment Group 2 (IDs)	-0.0115	-0.0184	-0.4541	-0.0821	0.0029	0.0132	-0.0100	-0.0039
	(0.0583)	(0.0578)	(0.5347)	(0.6047)	(0.0104)	(0.0105)	(0.0155)	(0.0221)
Constant	0.8897***	1.7307***	7.8228***	11.4313***	0.0981***	0.2907***	0.2500***	-0.0896*
	(0.0426)	(0.1432)	(0.3693)	(1.6370)	(0.0072)	(0.0363)	(0.0098)	(0.0520)
Controls	N	Y	N	Y	N	Y	N	Y
Observations	60,002	60,002	60,002	60,002	60,002	60,002	60,002	60,002
# Drivers	3,065	3,065	3,065	3,065	3,065	3,065	3,065	3,065
# Sites (clusters)	82	82	82	82	82	82	82	82
Adjusted R-squared	0.001	0.125	0.002	0.049	0.006	0.055	0.001	0.100

For all dependent variables, lower scores represent better performance (more efficient driving). This table reproduces the models of Table 3 on the matched sample. Errors clustered by site. ***, **, * indicate significance at 1%, 5% and 10%.

Table A4: Table 4 Post-Treatment Only

Dependent variable:	Log(Gap Score)		Shift Score		Log(Idle Time)		Log(Fuel Lost)	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Treatment Group 1*Phase 1	0.2896***	0.1616**	1.3842	1.0252	0.0489**	0.0175	0.0762***	0.0553**
	(0.0903)	(0.0636)	(1.0773)	(0.8802)	(0.0221)	(0.0186)	(0.0233)	(0.0243)
Treatment Group 2*Phase 1	0.0817	0.1218	1.3051	0.9836	-0.0066	0.0228	0.0075	0.0413
	(0.1017)	(0.0787)	(1.0149)	(0.9623)	(0.0269)	(0.0176)	(0.0276)	(0.0312)
Treatment Group 1 (names)	-0.1150**	-0.0626*	-0.1022	0.2204	-0.0231*	-0.0197*	-0.0223	-0.0209
	(0.0568)	(0.0355)	(0.6724)	(0.5670)	(0.0127)	(0.0104)	(0.0152)	(0.0134)
Treatment Group 2 (IDs)	-0.0636	-0.0240	-0.9529*	-0.0756	0.0205	0.0116	-0.0035	-0.0120
	(0.0699)	(0.0449)	(0.5481)	(0.7222)	(0.0204)	(0.0151)	(0.0156)	(0.0177)
Phase 1	-0.1333**	-0.1208***	-1.0758	-1.1257*	0.0029	-0.0242**	-0.0243	-0.0371**
	(0.0601)	(0.0426)	(0.6786)	(0.6775)	(0.0148)	(0.0119)	(0.0160)	(0.0166)
Constant	0.9416***	1.8361	7.4250***	8.0209	0.1181***	0.2978	0.2623***	-0.0538
	(0.0445)	(0.1502)	(0.3639)	(1.8019)	(0.0098)	(0.0460)	(0.0100)	(0.0553)
Controls	N	Y	N	Y	N	Y	N	Y
Observations	46,848	46,848	46,848	46,848	46,848	46,848	46,848	46,848
# Drivers	4,597	4,597	4,597	4,597	4,597	4,597	4,597	4,597
# Sites (clusters)	143	143	143	143	143	143	143	143
Adjusted R-squared	0.024	0.146	0.010	0.050	0.024	0.066	0.011	0.114

For all dependent variables, lower scores represent better performance (more efficient driving). This table reproduces the models of Table 4 on the post-treatment observations only. Errors clustered by site. ***, **, * indicate significance at 1%, 5% and 10%.

Table A5: Placebo Test

	Actual MPG		Potential MPG	
	(1)	(2)	(3)	(4)
Post*Treatment Group 1*Phase 1	-0.0363***	-0.0223**	0.0158	-0.0560
	(0.0106)	(0.0100)	(0.0731)	(0.0527)
Post*Treatment Group 2*Phase 1	-0.0059	-0.0163	0.0981	0.0679
	(0.0141)	(0.0107)	(0.0777)	(0.0541)
Post*Treatment Group 1	0.0083	0.0082	0.0539	0.0589*
	(0.0071)	(0.0062)	(0.0495)	(0.0339)
Post*Treatment Group 2	-0.0109	-0.0024	-0.0408	-0.0353
	(0.0114)	(0.0082)	(0.0562)	(0.0348)
Post*Phase 1	0.0068	0.0081	-0.1213**	-0.0579
	(0.0088)	(0.0074)	(0.0481)	(0.0383)
Treatment Group 1*Phase 1	-0.0465*		-0.4144*	
	(0.0239)		(0.2316)	
Treatment Group 2*Phase 1	-0.0123		-0.3375	
	(0.0242)		(0.2204)	
Post	-0.0026	0.0056	-0.0798**	0.0289
	(0.0060)	(0.0059)	(0.0345)	(0.0373)
Phase 1	0.0238		0.1512	
	(0.0158)		(0.1694)	
Treatment Group 1 (names)	0.0168		0.1899	
	(0.0153)		(0.1493)	
Treatment Group 2 (IDs)	0.0235		0.0918	
	(0.0186)		(0.1432)	
Constant	0.0212	-0.0470***	6.8736***	6.7390***
	(0.0197)	(0.0112)	(0.1061)	(0.0264)
Date and Driver FE	N	Y	N	Y
Controls	Y	Y	Y	Y
Observations	93,913	93,913	93,913	93,913
# Drivers	4,827	4,827	4,827	4,827
# Sites (clusters)	143	143	143	143
Adjusted R-squared	0.973	0.985	0.159	0.577

Higher MPG represents better performance (more efficient driving). See Table 1 for variable definitions and Table 3 for description of basic specification. *Actual MPG* refers the recorded miles per gallon actually achieved on a given route. *Potential MPG* refers the the EOBR-calculated feasible miles per gallon for the same route. *Potential MPG* is system-calculated based on route and weather characteristics and is not intended to be affected by a driver's actions. Errors clustered by site. ***, **, * indicate significance at 1%, 5% and 10%.

Table A6: Placebo Test on Matched Sample

	Actual MPG		Potential MPG	
	(1)	(2)	(3)	(4)
Post*Treatment Group 1*Phase 1	-0.0357***	-0.0202**	0.0368	-0.0272
	(0.0127)	(0.0097)	(0.1032)	(0.0695)
Post*Treatment Group 2*Phase 1	-0.0015	-0.0050	0.1273	0.1413**
	(0.0160)	(0.0115)	(0.0945)	(0.0608)
Post*Treatment Group 1	0.0117	0.0096	-0.0094	0.0595
	(0.0092)	(0.0078)	(0.0780)	(0.0499)
Post*Treatment Group 2	-0.0093	-0.0081	-0.1068	-0.0812*
	(0.0138)	(0.0105)	(0.0725)	(0.0420)
Post*Phase 1	0.0079	0.0035	-0.0976	-0.0846*
	(0.0098)	(0.0061)	(0.0665)	(0.0464)
Treatment Group 1*Phase 1	-0.0322		-0.1821	
	(0.0314)		(0.2904)	
Treatment Group 2*Phase 1	0.0209		-0.0944	
	(0.0307)		(0.2387)	
Post	-0.0075		-0.0749	
	(0.0077)		(0.0513)	
Phase 1	0.0039		0.0043	
	(0.0224)		(0.1728)	
Treatment Group 1 (names)	0.0107		0.3404	
	(0.0200)		(0.2122)	
Treatment Group 2 (IDs)	-0.0118		0.1995	
	(0.0236)		(0.1770)	
Constant	0.0252	-0.0518***	6.7012***	6.6435***
	(0.0248)	(0.0136)	(0.1284)	(0.0346)
Date and Driver FE	N	Y	N	Y
Controls	Y	Y	Y	Y
Observations	60,002	60,002	60,002	60,002
# Drivers	3,065	3,065	3,065	3,065
# Sites (clusters)	82	82	82	82
Adjusted R-squared	0.973	0.985	0.159	0.571

Higher MPG represents more efficient driving. This table reproduces the models of Table A3 on the matched sample. Errors clustered by site. ***, **, * indicate significance at 1%, 5% and 10%.

Table A7: Instrumental Variables Analysis

Dependent variable:	Log(Gap Score)		Shift Score		Log(Idle Time)		Log(Fuel Lost)	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Post*Treatment Group 1*Phase 1	0.2543***	0.3043***	4.3828**	4.8073**	0.0843***	0.0785**	0.1228***	0.1228***
	(0.0898)	(0.1169)	(2.1067)	(2.2551)	(0.0293)	(0.0322)	(0.0443)	(0.0417)
Post*Treatment Group 2*Phase 1	0.0256	0.0327	0.7487	1.0842	-0.0298	-0.0300	0.0112	0.0185
	(0.1150)	(0.1119)	(1.6213)	(1.5123)	(0.0431)	(0.0408)	(0.0372)	(0.0375)
Post*Treatment Group 1	-0.0824*	-0.0717*	-0.2826	-0.1439	-0.0256*	-0.0230*	-0.0385***	-0.0302**
	(0.0460)	(0.0410)	(0.7190)	(0.6747)	(0.0136)	(0.0135)	(0.0143)	(0.0142)
Post*Treatment Group 2	0.0544	0.0776	0.6518	0.6611	0.0378	0.0348	0.0157	0.0147
	(0.0954)	(0.0925)	(1.1293)	(1.0624)	(0.0400)	(0.0377)	(0.0297)	(0.0311)
Post*Phase 1	-0.0394	-0.0330	-0.9913	-0.9974*	0.0021	0.0020	-0.0127	-0.0108
	(0.0340)	(0.0301)	(0.6310)	(0.5756)	(0.0082)	(0.0082)	(0.0117)	(0.0101)
Treatment Group 1*Phase 1	0.3463*	0.0902	-1.2734	-1.8784	0.0149	-0.0449	0.0401	0.0127
	(0.2095)	(0.1614)	(2.2568)	(2.0030)	(0.0372)	(0.0427)	(0.0553)	(0.0590)
Treatment Group 2*Phase 1	0.1748	0.2504	2.3213	1.3782	0.0026	0.0542	0.0052	0.0850
	(0.1792)	(0.1985)	(2.1894)	(1.9799)	(0.0375)	(0.0395)	(0.0524)	(0.0749)
Post	0.0148	0.0096	-0.1545	0.0466	0.0001	-0.0041	0.0058	0.0066
	(0.0355)	(0.0291)	(0.4271)	(0.3754)	(0.0089)	(0.0086)	(0.0108)	(0.0106)
Phase 1	-0.0934	-0.0953*	-0.0738	-0.3020	0.0005	-0.0186*	-0.0119	-0.0297*
	(0.0573)	(0.0487)	(0.6960)	(0.6508)	(0.0111)	(0.0102)	(0.0157)	(0.0180)
Treatment Group 1 (names)	-0.1133	-0.0614	0.1504	0.4500	-0.0168	-0.0121	0.0005	-0.0149
	(0.0932)	(0.0674)	(1.1827)	(0.9834)	(0.0176)	(0.0199)	(0.0246)	(0.0260)
Treatment Group 2 (IDs)	-0.2330*	-0.1626	-3.1412*	-0.9666	0.0111	-0.0087	-0.0282	-0.0547
	(0.1368)	(0.1046)	(1.6121)	(1.3798)	(0.0290)	(0.0297)	(0.0380)	(0.0391)
Constant	0.9838***	1.5828***	8.0867***	3.5098**	0.1178***	0.1997***	0.2686***	-0.2450***
	(0.0454)	(0.1433)	(0.4608)	(1.4679)	(0.0103)	(0.0331)	(0.0120)	(0.0495)
Controls	N	Y	N	Y	N	Y	N	Y
Observations	93,913	93,913	93,913	93,913	93,913	93,913	93,913	93,913
# Drivers	4,827	4,827	4,827	4,827	4,827	4,827	4,827	4,827
# Sites (clusters)	143	143	143	143	143	143	143	143
Adjusted R-squared	0.024	0.128	0.003	0.044	0.019	0.055	0.008	0.104

For all dependent variables, lower scores represent better performance (more efficient driving). This table reproduces Table 4, but instruments the actual treatment of a given site by the assignment treatment. Errors clustered by site. ***, **, * indicate significance at 1%, 5% and 10%.

Table A8: Instrumental Variable Analysis on Matched Sample

Dependent variable:	Log(Gap Score)		Shift Score		Log(Idle Time)		Log(Fuel Lost)	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Post*Treatment Group 1*Phase 1	0.2478**	0.2779***	-0.3197	0.2788	0.0891***	0.0928***	0.1085***	0.1088***
	(0.1023)	(0.0962)	(1.7286)	(1.7160)	(0.0306)	(0.0327)	(0.0378)	(0.0328)
Post*Treatment Group 2*Phase 1	-0.1392	-0.0653	-2.2333	-1.5546	-0.1417	-0.1063	-0.0225	-0.0103
	(0.2786)	(0.2287)	(2.6756)	(2.5078)	(0.1317)	(0.1075)	(0.0776)	(0.0771)
Post*Treatment Group 1	-0.1258*	-0.1138**	1.5508	1.1777	-0.0441**	-0.0464**	-0.0465**	-0.0445**
	(0.0664)	(0.0550)	(1.2131)	(1.2117)	(0.0196)	(0.0198)	(0.0195)	(0.0185)
Post*Treatment Group 2	0.1708	0.1316	2.0884	1.6287	0.1386	0.1053	0.0366	0.0271
	(0.2691)	(0.2195)	(2.5385)	(2.3584)	(0.1283)	(0.1055)	(0.0755)	(0.0735)
Post*Phase 1	-0.0446	-0.0334	0.2338	0.2308	-0.0021	-0.0039	-0.0118	-0.0089
	(0.0375)	(0.0315)	(0.4689)	(0.4510)	(0.0103)	(0.0100)	(0.0113)	(0.0116)
Treatment Group 1*Phase 1	0.1737	0.1729	-0.2868	2.3143	0.0208	0.0231	-0.0158	0.0749
	(0.2315)	(0.2374)	(2.0532)	(2.3838)	(0.0405)	(0.0509)	(0.0545)	(0.0867)
Treatment Group 2*Phase 1	-0.2368	0.4591	2.6560	6.3768*	-0.0321	0.1318	-0.0864	0.1960
	(0.4042)	(0.3189)	(2.8850)	(3.4571)	(0.0718)	(0.0836)	(0.0976)	(0.1263)
Post	0.0312	0.0241	-1.0002**	-0.3598	0.0078	0.0063	0.0079	0.0113
	(0.0465)	(0.0401)	(0.4882)	(0.4623)	(0.0128)	(0.0130)	(0.0136)	(0.0158)
Phase 1	-0.0007	-0.0774	0.2917	-1.1688	-0.0024	-0.0261	0.0157	-0.0384
	(0.0774)	(0.0807)	(0.7204)	(0.9145)	(0.0117)	(0.0166)	(0.0184)	(0.0310)
Treatment Group 1 (names)	-0.0500	0.0099	-0.1087	-0.2761	-0.0050	-0.0179	0.0038	0.0124
	(0.1145)	(0.1477)	(1.1666)	(1.5567)	(0.0247)	(0.0432)	(0.0284)	(0.0582)
Treatment Group 2 (IDs)	0.1929	-0.3043*	-2.9485	-4.1722*	0.0469	-0.0676	0.0464	-0.1152*
	(0.3789)	(0.1701)	(2.4942)	(2.2000)	(0.0690)	(0.0588)	(0.0898)	(0.0681)
Constant	0.9081***	1.9561***	7.3256***	36.4616***	0.1194***	0.2464***	0.2632***	0.0341
	(0.0577)	(0.1619)	(0.5156)	(1.6168)	(0.0125)	(0.0430)	(0.0158)	(0.0583)
Controls	N	Y	N	Y	N	Y	N	Y
Observations	60,002	60,002	60,002	60,002	60,002	60,002	60,002	60,002
# Drivers	3,065	3,065	3,065	3,065	3,065	3,065	3,065	3,065
# Sites (clusters)	82	82	82	82	82	82	82	82
Adjusted R-squared	0.003	0.120	0.009	0.047	0.008	0.058	0.009	0.091

For all dependent variables, lower scores represent better performance (more efficient driving). This table reproduces the analysis in Table A5 on the matched sample. Errors clustered by site. ***, **, * indicate significance at 1%, 5% and 10%.

Table A9: Effect of Rankings on Phase 1 and Pre-Phase 1 Sites - Long Window

Dependent variable:	Log(Gap Score)		Shift Score		Log(Idle Time)		Log(Fuel Lost)	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Post*Treatment Group 1*Phase 1	0.0966**	0.0963**	1.6616	1.4921	0.0259	0.0252	0.0465***	0.0373***
	(0.0441)	(0.0429)	(1.0370)	(1.0080)	(0.0163)	(0.0156)	(0.0156)	(0.0140)
Post*Treatment Group 2*Phase 1	0.0263	0.0297	1.5130	1.4498	-0.0036	0.0001	0.0130	0.0131
	(0.0486)	(0.0464)	(0.9732)	(0.9436)	(0.0183)	(0.0167)	(0.0185)	(0.0161)
Post*Treatment Group 1	-0.0414	-0.0526**	-0.6129	-0.6155	-0.0116	-0.0114	-0.0259***	-0.0201**
	(0.0270)	(0.0248)	(0.6369)	(0.6179)	(0.0082)	(0.0073)	(0.0093)	(0.0090)
Post*Treatment Group 2	0.0289	0.0295	0.0243	0.0978	0.0168	0.0141	0.0065	0.0069
	(0.0354)	(0.0325)	(0.7387)	(0.7135)	(0.0127)	(0.0117)	(0.0133)	(0.0113)
Post*Phase 1	-0.0280	-0.0332	-0.3156	-0.3168	0.0066	0.0023	-0.0124	-0.0115
	(0.0328)	(0.0317)	(0.7552)	(0.7297)	(0.0121)	(0.0111)	(0.0117)	(0.0103)
Treatment Group 1*Phase 1	0.1407	0.0377	-0.3884	-0.9014	0.0100	-0.0046	0.0105	0.0076
	(0.0870)	(0.0675)	(1.3413)	(1.1558)	(0.0168)	(0.0161)	(0.0227)	(0.0246)
Treatment Group 2*Phase 1	0.0653	0.0556	0.1648	-0.8462	-0.0000	0.0080	0.0007	0.0194
	(0.0852)	(0.0692)	(1.1586)	(1.0033)	(0.0145)	(0.0157)	(0.0230)	(0.0269)
Post	0.0471**	0.0532***	-1.6888***	-1.6883***	0.0170***	0.0189***	0.0229***	0.0198***
	(0.0196)	(0.0167)	(0.4700)	(0.4476)	(0.0060)	(0.0053)	(0.0065)	(0.0058)
Phase 1	-0.0936*	-0.0628	-0.7248	-0.3364	-0.0034	-0.0166	-0.0121	-0.0200
	(0.0561)	(0.0487)	(0.8861)	(0.7853)	(0.0096)	(0.0107)	(0.0148)	(0.0180)
Treatment Group 1 (names)	-0.0499	-0.0072	0.0830	0.4096	-0.0081	-0.0076	0.0038	0.0007
	(0.0553)	(0.0325)	(0.9339)	(0.7730)	(0.0101)	(0.0095)	(0.0131)	(0.0124)
Treatment Group 2 (IDs)	-0.1021*	-0.0355	-1.4287*	-0.2472	-0.0011	0.0031	-0.0164	-0.0121
	(0.0614)	(0.0438)	(0.8018)	(0.7815)	(0.0104)	(0.0109)	(0.0142)	(0.0170)
Constant	0.9776***	1.7193***	10.6972***	13.5131***	0.0953***	0.2802***	0.2588***	-0.1257***
	(0.0428)	(0.1134)	(0.5702)	(1.3161)	(0.0071)	(0.0289)	(0.0085)	(0.0440)
Controls	N	Y	N	Y	N	Y	N	Y
Observations	310,084	310,084	310,084	310,084	310,084	310,084	310,084	310,084
# Drivers	5,011	5,011	5,011	5,011	5,011	5,011	5,011	5,011
# Sites (clusters)	143	143	143	143	143	143	143	143
Adjusted R-squared	0.008	0.128	0.010	0.052	0.006	0.048	0.004	0.107

For all dependent variables, lower scores represent better performance (more efficient driving). This table reproduces the analysis in Table 4 on the full window of data (inclusive of 47 days prior to posting through 207 days post), exclusive of the five days around the posting window. Note that the data collection period ended 85 days after the last posting date, so there is attrition in the sample during the later dates, with only the earliest 11 sites in the experiment having data more than 200 days after posting. Errors clustered by site. ***, **, * indicate significance at 1%, 5% and 10%.

Table A10: Driver Fixed Effects - Long Window

Dependent variable:	Log(Gap Score)		Shift Score		Log(Idle Time)		Log(Fuel Lost)	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Post*Treatment Group 1*Phase 1	0.0829*	0.0822*	1.4671	1.429	0.0264*	0.0228	0.0375**	0.0322**
	(0.0423)	(0.0426)	(1.0750)	(1.0707)	(0.0154)	(0.0148)	(0.0154)	(0.0144)
Post*Treatment Group 2*Phase 1	0.035	0.0333	1.6876*	1.6902*	-0.0026	-0.0013	0.0127	0.0154
	(0.0399)	(0.0402)	(1.0178)	(1.0144)	(0.0152)	(0.0148)	(0.0147)	(0.0139)
Post*Treatment Group 1	-0.0516**	-0.0534**	-0.7238	-0.7138	-0.0095	-0.0074	-0.0251**	-0.0209**
	(0.0259)	(0.0254)	(0.6759)	(0.6738)	(0.0070)	(0.0069)	(0.0097)	(0.0097)
Post*Treatment Group 2	0.0095	0.0111	0.0248	0.0191	0.0123	0.0107	0.0037	0.0005
	(0.0281)	(0.0276)	(0.7814)	(0.7782)	(0.0108)	(0.0105)	(0.0098)	(0.0096)
Post*Phase 1	-0.031	-0.0319	-0.4345	-0.4371	0.002	0.0023	-0.0128	-0.0119
	(0.0283)	(0.0288)	(0.7643)	(0.7586)	(0.0097)	(0.0093)	(0.0104)	(0.0096)
Constant	0.9637***	1.8205***	9.5121***	11.0308***	0.1193***	0.1670***	0.2628***	-0.0866***
	(0.0126)	(0.0411)	(0.3789)	(0.6108)	(0.0057)	(0.0105)	(0.0054)	(0.0184)
Controls	N	Y	N	Y	N	Y	N	Y
Observations	310,084	310,084	310,084	310,084	310,084	310,084	310,084	310,084
# Drivers	5,011	5,011	5,011	5,011	5,011	5,011	5,011	5,011
# Sites (clusters)	143	143	143	143	143	143	143	143
Adjusted R-squared	0.515	0.527	0.526	0.529	0.264	0.277	0.430	0.483

For all dependent variables, lower scores represent better performance (more efficient driving). This table reproduces the analysis in Table A7 but with driver fixed effects included. Errors clustered by site. ***, **, * indicate significance at 1%, 5% and 10%.

Table A11: Matched Analysis - Long Window

Dependent variable:	Log(Gap Score)		Shift Score		Log(Idle Time)		Log(Fuel Lost)	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Post*Treatment Group 1*Phase 1	0.1230**	0.1366**	-0.6115	-0.5885	0.0573***	0.0536***	0.0645***	0.0528***
	(0.0570)	(0.0522)	(1.4459)	(1.3966)	(0.0186)	(0.0176)	(0.0196)	(0.0187)
Post*Treatment Group 2*Phase 1	0.0116	0.0290	-0.9394	-0.9843	-0.0025	0.0012	0.0125	0.0097
	(0.0582)	(0.0550)	(1.4392)	(1.3895)	(0.0188)	(0.0172)	(0.0227)	(0.0199)
Post*Treatment Group 1	-0.0563	-0.0702**	-0.2320	-0.1899	-0.0238*	-0.0228*	-0.0333***	-0.0281**
	(0.0365)	(0.0324)	(0.7581)	(0.7241)	(0.0129)	(0.0116)	(0.0123)	(0.0118)
Post*Treatment Group 2	0.0429	0.0304	-0.1244	-0.1089	0.0216	0.0176	0.0063	0.0070
	(0.0452)	(0.0429)	(1.0250)	(0.9792)	(0.0168)	(0.0156)	(0.0178)	(0.0150)
Post*Phase 1	-0.0524	-0.0629*	-0.2213	-0.1467	-0.0148	-0.0152	-0.0250	-0.0197
	(0.0412)	(0.0370)	(1.1712)	(1.1342)	(0.0105)	(0.0091)	(0.0151)	(0.0141)
Treatment Group 1*Phase 1	0.0791	0.0207	0.0990	-0.6190	0.0041	0.0015	-0.0115	0.0110
	(0.1113)	(0.0900)	(1.5884)	(1.4769)	(0.0229)	(0.0221)	(0.0247)	(0.0317)
Treatment Group 2*Phase 1	-0.0633	0.0739	0.9075	-0.0142	-0.0068	0.0076	-0.0311	0.0371
	(0.1058)	(0.0798)	(1.5747)	(1.3542)	(0.0176)	(0.0160)	(0.0265)	(0.0301)
Post	0.0626**	0.0735***	1.7184***	1.6373***	0.0265***	0.0261***	0.0308***	0.0266***
	(0.0277)	(0.0229)	(0.6430)	(0.6093)	(0.0087)	(0.0079)	(0.0093)	(0.0083)
Phase 1	0.0007	-0.0004	-0.3159	0.3830	-0.0024	-0.0073	0.0173	-0.0074
	(0.0786)	(0.0651)	(1.2546)	(1.1134)	(0.0124)	(0.0114)	(0.0168)	(0.0246)
Treatment Group 1 (names)	-0.0248	0.0284	0.7930	0.7404	-0.0028	-0.0092	0.0014	0.0145
	(0.0701)	(0.0604)	(1.0287)	(0.9204)	(0.0149)	(0.0165)	(0.0153)	(0.0231)
Treatment Group 2 (IDs)	0.0266	-0.0850	0.7728	1.0107	0.0050	-0.0025	0.0024	-0.0339
	(0.0761)	(0.0630)	(1.0768)	(1.1107)	(0.0126)	(0.0126)	(0.0181)	(0.0240)
Constant	0.8986***	1.5815***	91.0687***	88.5450***	0.0929***	0.2807***	0.2459***	-0.1567***
	(0.0562)	(0.1392)	(0.7216)	(1.5272)	(0.0086)	(0.0313)	(0.0106)	(0.0503)
Controls	N	Y	N	Y	N	Y	N	Y
Observations	198,831	198,831	198,831	198,831	198,831	198,831	198,831	198,831
# Drivers	3,186	3,186	3,186	3,186	3,186	3,186	3,186	3,186
# Sites (clusters)	82	82	82	82	82	82	82	82
Adjusted R-squared	0.008	0.130	0.006	0.051	0.008	0.051	0.005	0.108

For all dependent variables, lower scores represent better performance (more efficient driving). This table reproduces the analysis in Table A7 but on the matched sample. Errors clustered by site. ***, **, * indicate significance at 1%, 5% and 10%.

Table A12: Matched Analysis with Driver Fixed Effects - Long Window

Dependent variable:	Log(Gap Score)		Shift Score		Log(Idle Time)		Log(Fuel Lost)	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Post*Treatment Group 1*Phase 1	0.0990*	0.1004*	0.5058	0.4726	0.0445**	0.0389**	0.0464**	0.0378*
	(0.0549)	(0.0541)	(1.4009)	(1.3983)	(0.0177)	(0.0175)	(0.0199)	(0.0194)
Post*Treatment Group 2*Phase 1	0.0031	0.0099	1.1112	1.1595	-0.0131	-0.0117	0.004	0.0025
	(0.0466)	(0.0460)	(1.4296)	(1.4241)	(0.0156)	(0.0156)	(0.0165)	(0.0166)
Post*Treatment Group 1	-0.0566	-0.0582	0.2796	0.2884	-0.0203**	-0.0179*	-0.0290**	-0.0249*
	(0.0361)	(0.0354)	(0.7970)	(0.7936)	(0.0093)	(0.0091)	(0.0137)	(0.0135)
Post*Treatment Group 2	0.0319	0.0299	0.2026	0.1741	0.0198	0.0174	0.0095	0.0071
	(0.0359)	(0.0353)	(1.0579)	(1.0549)	(0.0134)	(0.0134)	(0.0124)	(0.0123)
Post*Phase 1	-0.0375	-0.0439	-0.0202	-0.0464	-0.0054	-0.0041	-0.0176	-0.0126
	(0.0354)	(0.0341)	(1.1392)	(1.1336)	(0.0082)	(0.0084)	(0.0128)	(0.0130)
Constant	0.9436***	1.7822***	9.1543***	22.5564***	0.1204***	0.1765***	0.2631***	-0.0862***
	(0.0167)	(0.0558)	(0.5084)	(0.7706)	(0.0074)	(0.0129)	(0.0069)	(0.0233)
Controls	N	Y	N	Y	N	Y	N	Y
Observations	198,831	198,831	198,831	198,831	198,831	198,831	198,831	198,831
# Drivers	3,186	3,186	3,186	3,186	3,186	3,186	3,186	3,186
# Sites (clusters)	82	82	82	82	82	82	82	82
Adjusted R-squared	0.515	0.527	0.526	0.529	0.264	0.277	0.430	0.483

For all dependent variables, lower scores represent better performance (more efficient driving). This table reproduces the analysis in Table A8 but on the matched sample. Errors clustered by site. ***, **, * indicate significance at 1%, 5% and 10%.

Table A13: Phase 1 and Employee Engagement

Dependent variable:	Collective index		Instrumental index	
	(1)	(2)	(3)	(4)
Phase 1	0.2268	0.2735**	0.2954*	0.3314**
	(0.1606)	(0.1234)	(0.1753)	(0.1607)
Constant	3.3001***	3.2945***	3.2542***	2.7641***
	(0.1681)	(0.4946)	(0.1697)	(0.4358)
Demographic Controls	No	Yes	No	Yes
Observations	561	561	564	564
Adjusted R-squared	0.026	0.127	0.016	0.075

This table is constructed from an employee engagement survey conducted on a subset of sites. Dependent variables are two indices constructed distinct questions on the survey. *Collective index* is the average of those questions that capture the degree to which the employee identifies with the larger organization. *Instrumental index* is the average of a subset of questions that capture satisfaction with the formal aspects of the job, such as benefits and compensation. Demographic controls include *race*, *age* and *tenure at company*. Errors clustered by site. ***, **, * indicate significance at 1%, 5% and 10%.

Table A14: Matched Analysis of Phase 1 and Employee Engagement

Dependent variable:	Collective index		Instrumental index	
	(1)	(2)	(3)	(4)
Phase 1	0.1731	0.2736**	0.1730	0.2353
	(0.1577)	(0.1298)	(0.1978)	(0.1965)
Constant	3.3550***	3.3327***	3.3138***	2.9711***
	(0.2183)	(0.3734)	(0.1881)	(0.6146)
Demographic Controls	No	Yes	No	Yes
Observations	396	396	399	399
Adjusted R-squared	0.029	0.128	0.009	0.056

See Table A13 for a description of data and controls. This table replicates the analysis of Table 10 on a matched subsample of Phase 1 and pre-Phase 1 sites so that these sites matched on observable characteristics. Errors clustered by site. ***, **, * indicate significance at 1%, 5% and 10%.

Table A15: Collective values sub-categories

Dependent variable:	Log (Gap Score)							
Category:	Trust	Trust	Team	Team	Feel valued	Feel valued	Pride	Pride
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Post*Posting group 1*[Category]	0.1253***	0.1297**	0.0643*	0.0667	0.1123***	0.1079**	0.0544*	0.0621
	(0.0428)	(0.0553)	(0.0356)	(0.0408)	(0.0377)	(0.0402)	(0.0278)	(0.0378)
Post*Posting group 2*[Category]	0.0714	0.0612	-0.0405	-0.0820**	0.0204	-0.0273	-0.0091	-0.0177
	(0.0653)	(0.0858)	(0.0392)	(0.0395)	(0.0469)	(0.0503)	(0.0385)	(0.0527)
Post*Posting group 1	-0.4402***	-0.4783***	-0.2507**	-0.2840**	-0.3859***	-0.3958***	-0.2223**	-0.2731**
	(0.1296)	(0.1644)	(0.1086)	(0.1219)	(0.1073)	(0.1068)	(0.0908)	(0.1178)
Post*Posting group 2	-0.1710	-0.2293	0.1780	0.2225	-0.0076	0.0556	0.0767	0.0226
	(0.2328)	(0.3298)	(0.1400)	(0.1701)	(0.1714)	(0.2082)	(0.1640)	(0.2323)
Post*[Category]	-0.0736**	-0.0717	-0.0162	-0.0086	-0.0597**	-0.0429	-0.0338	-0.0313
	(0.0356)	(0.0488)	(0.0269)	(0.0317)	(0.0290)	(0.0333)	(0.0215)	(0.0328)
Posting group 1*[Category]	0.1063*	0.1206*	0.0776*	0.0739	0.0798	0.1001	0.1094*	0.1490***
	(0.0551)	(0.0633)	(0.0442)	(0.0522)	(0.0556)	(0.0598)	(0.0545)	(0.0527)
Posting group 2*[Category]	-0.0832	-0.0475	0.0329	0.1253*	0.0037	0.1382*	-0.0617	-0.0103
	(0.0568)	(0.0813)	(0.0712)	(0.0644)	(0.0767)	(0.0687)	(0.0528)	(0.0594)
Post	0.2924***	0.3035**	0.1094	0.1042	0.2393***	0.2046**	0.1716**	0.1797*
	(0.1078)	(0.1422)	(0.0792)	(0.0915)	(0.0827)	(0.0876)	(0.0726)	(0.0990)
[Category]	-0.0457	-0.0620	-0.0372	-0.0372	-0.0094	-0.0330	-0.0165	-0.0354
	(0.0360)	(0.0396)	(0.0289)	(0.0296)	(0.0396)	(0.0373)	(0.0339)	(0.0368)
Posting group 1 (Names)	-0.2910	-0.2991	-0.1903	-0.1251	-0.1936	-0.2104	-0.3190	-0.4095**
	(0.2011)	(0.2240)	(0.1757)	(0.1961)	(0.2012)	(0.2143)	(0.2129)	(0.1885)
Posting group 2 (IDs)	0.3544*	0.2715	-0.0033	-0.2609	0.0861	-0.3069	0.2802	0.1532
	(0.1851)	(0.2691)	(0.2213)	(0.1924)	(0.2481)	(0.2246)	(0.1726)	(0.1826)
Constant	2.1385***	2.2673***	2.1134***	2.1883***	1.9967***	2.1938***	2.0225***	2.1920***
	(0.2957)	(0.3322)	(0.3076)	(0.3571)	(0.3275)	(0.3737)	(0.3069)	(0.3283)
Sample	Full	Matched	Full	Matched	Full	Matched	Full	Matched
Observations	35,187	26,065	35,187	26,065	35,187	26,065	35,187	26,065
# Drivers	491	368	491	368	491	368	491	368
# Sites (clusters)	43	26	43	26	43	26	43	26
Adjusted R-squared	0.105	0.114	0.094	0.103	0.099	0.113	0.105	0.121

This table replicates the analysis of Table 10, disaggregating the "Collective" index into its constituent subindices. Errors clustered by site. ***, **, * indicate significance at 1%, 5% and 10%.