Does Targeting Relative Performance Feedback Improve Worker Productivity? Field Experimental Evidence from Bus Drivers

Gert-Jan Romensen*
University of Groningen

Adriaan R. Soetevent†
University of Groningen

February 14, 2025

Abstract

An often-voiced concern with relative performance feedback is that it may not improve workplace productivity if workers become demotivated and see no way to improve. Targeting feedback at specific productivity measures over which workers have direct control may in such cases prevent demotivation and focus attention. Does targeting improve worker productivity? We partner with a large bus company and experimentally vary the nature and number of peer-comparison messages which 409 bus drivers receive in their monthly feedback report. Messages are targeted at concrete driving behaviors and aimed at improving comfort and fuel efficiency. Using over 800,000 trip-level observations, we find that these targeted peer-comparison messages do not improve aggregate (fuel economy) or disaggregate measures (such as acceleration) of driving behavior. Further analyses also reveal no temporal or heterogeneous effects of the targeted messages.

JEL classification: D23, J24, M53, Q55.
Keywords: labor productivity, feedback, peer comparisons, field experiment, two-way fixed effects.

*University of Groningen, Faculty of Economics and Business, Nettelbosje 2, 9747 AE Groningen, The Netherlands, g.j.romensen@rug.nl. We thank Viola Angelini, Robert Dur, Mikhail Freer, Marco Haan, Patrick Legros, John List, Erzo Luttmer, Erin Mansur, Noemi Pace, Noémi Péter, Peter Siminski and seminar participants at AFE 2016, BIOECON 2017, M-BEES & M-BEPS 2017, EEA 2017, IMEBESS 2017, ESA Vienna 2017, IAEE 2018, ESA Los Angeles 2019, EALE 2021, the workshop Recognition and Feedback at Erasmus University, the Arne Ryde Workshop, University of Adelaide, Universitat Autònoma de Barcelona, Dartmouth College, ECARES, Jiatong University Beijing, Maastricht University, Northeastern University, University of St.Gallen, and Ca' Foscari University of Venice for their valuable comments. We are especially grateful to Peter Boersma, Marten Feenstra and Wouter van der Meer of Arriva for their time, support and excellent assistance in enabling this project. Views and opinions expressed in this paper as well as all remaining errors are solely those of the authors. An earlier version of the paper has circulated under the title "Tailored Feedback and Worker Green Behavior". This study is registered in the AEA RCT Registry (#AEARCTR-0005391).

†Corresponding author. University of Groningen, EEF, P. O. Box 800, 9700 AV Groningen, The Netherlands, a.r.soetevent@rug.nl.

1 Introduction

Feedback in the workplace often incorporates peer comparisons to improve worker productivity. Worker responses, however, may be mixed and dependent on the relative rank of the recipient. A concern, for example, is that recipients with a poor relative rank become demotivated and exert less effort as a result (Ashraf et al. 2014). Existing research has therefore called for the customization of relative performance feedback (Kuhnen and Tymula 2012) and has pointed to digital monitoring technologies in the workplace (Staats et al. 2017) as a promising way to facilitate customization because they routinely measure several productivity dimensions at once. How feedback should be customized to effectively target productivity remains an open question.

We contribute to answering this question by conducting a field experiment among all 409 tenured bus drivers at a public transport company that is installing electronic on-board recorders (EOBRs) in its bus fleet. For every worker, EOBRs measure the trip-level aggregate productivity (fuel efficiency) as well as the underlying performance on acceleration, braking, and cornering (ABC). In this setting we evaluate tailored relative performance feedback by introducing treatment variation in feedback intensity. Each driver is assigned random combinations of x corrective, and y positive messages that target the underlying ABC driving dimensions. These peer-comparison messages are integrated in monthly individual feedback reports that the company introduces as part of a campaign to stimulate efficient and comfortable driving. We vary the dosage of corrective feedback or supplement it with positive messages to test whether this approach avoids discouragement and bolsters workers' motivation to increase effort

Studies on relative performance feedback commonly find that the response to rankings depends on the relative rank of the recipient. In a study among private car drivers, Choudhary et al. (2022) find that high-performance drivers respond best to nudges that urge them to beat their personal best, whereas low-performing drivers benefit most from personal average nudges. We reason that part of the heterogeneity may be because feedback is typically given on final outcomes rather than on behaviors that lead to these outcomes. Little guidance on where to improve may create feelings of not being in control. Workers may derive more motivation from relative rankings that are based on specific productivity dimensions that are under their control.

We find in our field experiment, however, that targeting peer-comparison messages has no impact on worker productivity. Varying the number and nature of feedback messages on relative performance does not lead to changes in driving behavior among bus drivers. This holds for both aggregate (fuel economy) and disaggregate (acceleration, braking and cornering) measures

of worker effort. Further analyses suggest that there are also no temporal- or heterogeneous effects from tailoring peer-comparison feedback. Overall, the findings highlight the challenge of improving worker productivity through the customization of relative performance feedback.

This paper proceeds as follows. Section 2 motivates the study and reviews related literature. Section 3 describes the feedback program and the research design. The empirical evaluation of the feedback program is in Section 4. Section 5 concludes.

2 Motivation and literature review

A large literature shows that management practices matter for worker productivity (Bloom and van Reenen 2007, Syverson 2011, Bloom et al. 2013). Yet despite a considerable body of empirical work on relative performance feedback, the question how rank information affects worker productivity has not yet received its definite answer. Previous studies indicate that relative performance feedback can improve worker productivity (Blanes i Vidal and Nossol 2011, Song et al. 2018, Rilke et al. 2022), sales growth (Delfgaauw et al. 2013) and student performance (Azmat and Iriberri 2010, Tran and Zeckhauser 2012). Other studies report decreased performance following the provision of rank information (Bandiera et al. 2013, Ashraf et al. 2014) and improved performance when this information is abolished (Barankay 2012).

The literature has identified several channels through which relative performance feedback can improve worker productivity. Relative income concerns (Clark and Oswald 1996, Luttmer 2005) may spur effort if compensation is based on performance. Peer pressure may increase effort when revealing workers' relative performance activates social and reputational concerns (Kandel and Lazear 1992, Falk and Ichino 2006, Mas and Moretti 2009). In our setting, however, feedback is shared privately and drivers receive a fixed wage. Yet relative performance feedback may still improve performance because rank can be an inherent incentive (Tran and Zeckhauser 2012) and influence reward-related brain activity directly (Dohmen et al. 2011). For example, people's self-esteem (Kuhnen and Tymula 2012) and intrinsic motivation (Ryan and Deci 2000) may depend on their relative standing and relatedness with peers. Feedback recipients may be *behind-averse* or *ahead-seeking* (Roels and Su 2014), which motivates them to achieve a high relative rank. Studies have also shown that merely announcing relative performance feedback induces workers to work harder (Blanes i Vidal and Nossol 2011, Kuhnen and Tymula 2012).

Relative performance feedback may also negatively impact worker productivity. Some studies

have documented drops in performance when workers learn about their poor relative performance (Bandiera et al. 2013, Barankay 2012). In such cases, workers might prefer instead a fuzzy signal that enables them to maintain optimistic beliefs about their relative ability (Ashraf et al. 2014) or feedback that strengthens their confidence in being able to improve (Fischer and Sliwka 2018, Fischer and Wagner 2023). Workers may also respond to low relative ranks by handicapping their performance in order to maintain a positive self-image (Benabou and Tirole 2002). Finally, contextual factors may cause relative performance feedback to have a negative impact on worker effort. Blader et al. (2020) find a negative effect of relative performance information among truck drivers in plants with a teamwork culture.

Rankings are typically reported on final outcomes rather than on the intermediate steps leading to these outcomes. This may explain part of the heterogeneity in findings. Aggregated messages on outcomes may be demotivating because they give little guidance on where to improve and signal that improvement requires one big step rather than several small and clear steps. Bandiera et al. (2015) argue that university students often have imperfect information on how their effort translates into exam scores. To strategically use relative performance feedback to improve employee performance, Kuhnen and Tymula (2012) suggest that it might be promising to customize relative performance feedback by tailoring the content or by targeting subsets of workers. London (2003, pp 15-16) argues that feedback is only constructive if it is "relevant to elements of performance that contribute to task success and that are under the recipient's control". Our design aims to do just that.

The feeling of being in control is a key source of human motivation (Ryan and Deci 2000). Based on the reasoning that workers will exert more effort when they have direct control over the targeted outcomes, we introduce experimental variation in the number of corrective and positive feedback messages that target disaggregate, concrete areas of improvement. This allows us to test which combinations avoid discouragement and bolster worker effort. We hypothesize that the addition of targeted peer-comparison messages to a general feedback report will, on average, improve driver performance. Using the same argument of providing workers more control over their performance, Rilke et al. (2022) evaluate input-based instead of output-based relative performance feedback. They observe increased effort provided by low performing workers.

3 Research design

3.1 Field setting

Our field partner in this study is Arriva, a large passenger transport operator in Europe. Bus transport is Arriva's largest business unit. Most bus drivers are tenured employees (with only 14% on temporary contracts) and perform few other tasks in the organization other than driving bus routes. These routes depart from one of the six base locations and are typically inter-city routes in rural areas. Only the largest location has a mix of urban and rural routes. Drivers rotate shifts every week within a location and are therefore familiar with all routes in the area. Each route is thus assigned to multiple drivers and experienced under various on-the-road conditions. Promotion opportunities are limited for bus drivers and institutional constraints hinder the use of pay-for-performance schemes for economical and comfortable driving.

As part of its EcoManager campaign to stimulate efficient and comfortable driving, Arriva has equipped its entire bus fleet with electronic on-board recorders (EOBRs) that measure a bus driver's trip-level performance on several dimensions of driving behavior. The system records fuel consumption as well as comfort dimensions such as acceleration, braking and cornering (ABC). To measure performance on the ABC dimensions, the company formulated minimum performance threshold levels based on test rides under various conditions. An 'event' is recorded by the EOBR whenever a driver's action violates a threshold. The performance metric of the ABC dimensions is the number of events per 10km, with fewer events indicating better driving. These outcome data are uniquely matched with driver identifiers and centralized databases with information on driver and trip characteristics.

3.2 Research setting

The EOBR data are used as input to monthly feedback reports that are distributed among the drivers. The company also implemented a coaching initiative wherein drivers receive constructive feedback and guidance from an experienced colleague during on-the-road sessions. This coaching program is discussed and evaluated in Soetevent and Romensen (2024). We join the implementation process of the feedback reports in one of Arriva's concession areas in the Netherlands, covering about two-thirds of the province of Friesland.[1] In 2015, the number of check-ins with a public transport card indicates that Arriva serves a little more than five million travelers

[1] In the Netherlands, companies are awarded bus concessions through a competitive tendering process.

in this region.

The timeline is as follows. The old on-board system in the company's bus fleet has baseline trip-level data on fuel consumption that goes back to January 2015. Baseline data on the ABC dimensions from the new EOBR system is available from September 2015 onwards. During this baseline period, drivers are not aware of the upcoming feedback program and that they are being monitored. On October 5, 2015, Arriva distributed promotion material about the feedback program in the different locations to raise awareness among the bus drivers. The official launch of the program was during a kickoff event on November 9, 2015. All drivers were notified during this event about the monitoring of driving behavior and the introduction of monthly personalized feedback reports from December 2015 onwards. Furthermore, at this date, a LED-array in the buses is switched on as well to provide all drivers with instant feedback while driving. To rule out career concerns, bus drivers were informed that none of the feedback will be used in formal evaluations. Apart from the mentioned feedback forms, the company did not employ other incentives during the sample period to stimulate economical driving.

The monthly feedback report that drivers receive from December 15 onwards contains a letter score, ranging from A (highest score) to D (lowest score), on the ABC dimensions and fuel economy.[2] These scores are determined relative to a pre-set benchmark set by the company, so that in principle all drivers can get the highest score if they put in the effort. The scores do not reveal to drivers how they perform relative to colleagues and are not very informative about the dimensions of driving behavior they can relatively improve the most on. The feedback therefore gives little guidance and may therefore not motivate much to change driving behavior.[3] Our field experiment aims to change this by providing the relevant information through targeted peer-comparison feedback messages in the general feedback report. These messages will be part of the report for eleven rounds. The peer-comparison feedback period ends on November 15, 2016, with a notification to the drivers that the messages are no longer included.[4] The period after mid-November 2016 will be referred to as the post-experimental period. Data are collected

[2]Figure A.1 in Online Appendix A shows a sample report.

[3]In Soetevent and Romensen (2024), we examine the overall impact of the announcement and introduction of the general feedback reports. Driver performance in the current concession area is compared with the performance of drivers in two concession areas in Southwest Netherlands. These other areas are located about 200km away from the treatment region and do not yet have feedback reports in place at the time of the announcement and introduction in Friesland. Using a difference-in-differences approach, we find no indication that either the announcement or the introduction of the feedback reports improves worker productivity.

[4]The text of this one-time notification via the feedback reports is as follows (translated from Dutch): "Dear colleague, starting this month, this report will no longer include information about your performance relative to your colleagues". All drivers assigned to a treatment condition with peer-comparison messages received this notification.

until January 31, 2017.

3.3 Field experiment: Targeted peer-comparison treatments

Within the company, there was *a priori* a strong conviction that the largest gains in performance would be achieved with corrective forms of peer-comparison feedback, emphasizing the dimensions on which the driver could improve. We however worried that providing corrective feedback on too many driving dimensions at once, and providing corrective feedback only, would cause feedback overload or demotivate employees. To test whether corrective feedback is more effective when dosed or combined with positive feedback we designed the treatment conditions summarized in Table 1. The peer-comparison messages provide drivers with information on the ABC-comfort dimensions (but not fuel economy) that is complementary to the general feedback report: the peer-comparison messages inform drivers how they rank against their peers rather than against an overall benchmark.

Treatment T1 [0n0p] is the control condition with no peer-comparison messages. This controls for the effects of the general feedback report.[5] In treatment T2 [1n0p], one negative message is provided if drivers underperform on a particular ABC-dimension. That is, they are explicitly informed that they rank poorly compared to peers and are encouraged to improve. In T3 [1n1p], drivers additionally have a chance of receiving one positive message. In this case, they are told their good ranking and are encouraged to keep up the good work. If a driver performs poor (or well) on multiple dimensions, one will be randomly chosen. Finally, in T4 [3n0p], drivers may receive corrective feedback on all comfort dimensions. For example, in T3 [1n1p] the precise text of the messages may read as follows:

> Dear colleague,
>
> In terms of taking corners, you belong to the top 25 percent of the bus drivers in your location.
> You are doing excellent on this dimension!
>
> In terms of braking, you belong to the bottom 50 percent of the bus drivers in your location.
> You can improve on this dimension!

We randomly assign the complete universe of all 409 tenured drivers to one of the four experimental conditions, stratified along the dimensions of base location, gender, and years of service

[5] A printed version of the general report is delivered around the 15th day of each feedback month via the team manager or pigeonhole. Drivers in the control condition receive the same feedback report but without peer comparison messages. Company policies prohibit the creation of a control group that does not receive any form of feedback. By handing out reports to drivers in the control condition, we embed the experimental variation more naturally and explicitly recognize and control for Hawthorne and general feedback effects.

at the company. We exclude the 67 drivers with a temporary contract because their behavior is only observed for short and irregular time spans. Power calculations show that with 102/103 drivers per treatment arm, our design is moderately powered and able to detect small effect sizes.[6] We construct reference groups in which driver performance on each comfort dimension is compared to colleagues with the same base location and treatment status.[7] This gives each driver a natural and homogeneous comparison group in which competition is likely to generate strong incentives (Lazear and Rosen 1981, Delfgaauw et al. 2013). The comfort dimensions are disaggregated measures of driving behavior over which drivers have a strong direct influence, thereby making the feedback as concrete and useful as possible to the recipients. At the start of each month, the company shares with us a summary of each driver's performance during the previous month. We use this information to assess how a driver performed compared to his/her peers and to assign peer-comparison messages.[8] Dependent on treatment assignment, a number of negative (positive) messages are provided if a driver belongs to the bottom 50% (top 25%) of the reference group. The peer-comparison messages are updated every month to inform about progress and to avoid drivers from slacking off.

At this point, it is important to stress that the treatments condition on the eligibility to receive negative (and positive) peer-comparison feedback but not on the actual exposure. For example, among individuals in treatment group T2 [1n0p], only drivers who score lower than half of their peers on at least one of the three comfort dimensions receive a corrective message in that feedback round. In case they perform poorly on multiple dimensions, one dimension is randomly selected for corrective feedback. In a given month, about 30% of the drivers in T2 performs well on all dimensions and therefore does not receive a message. Hence, the treatment effects that we present show the effect of treatment eligibility. They are conservative estimates of the effect of exposure to peer-comparison feedback, as only part of the group actually receives these messages in a given month. For each driving dimension, there is considerable month-to-

[6]The Standardized Effect Size (MDES) is 0.39, which implies that the minimum detectable effect size (MDE) for fuel economy is 0.40 liter/100km or 66 meter/liter (a 1.7% improvement from baseline) and for acceleration 0.875 events/10km (a 8.0% improvement).

[7]This is because pre-treatment information revealed that high and low scores are occasionally concentrated in base locations. Limiting peer-comparison groups to drivers with the same treatment status avoids indirect treatment interference (the relative ranking of drivers in one treatment automatically deteriorates when drivers in another condition experience a positive treatment effect).

[8]The performance summary contains information on the bus-specific percentile rank of the driver on each driving dimension (compared to all drivers in the concession area who also operated on that bus type in the previous month). The final percentile rank for each driving dimension is the sum of the percentile ranks of the driver on each bus type, weighted by the number of kilometers driven on that bus type in that month. Within a reference group, a driver's final percentile rank determines how (s)he has performed compared to his/her peers.

month variation in the group of drivers in the top-25% and bottom-50% group. Most drivers move in and out, but some drivers are never in the top (bottom) part.[9]

Table 2 summarizes per experimental condition the data in the final analysis sample and reports the outcome of balance tests. The p-values show that driver pre-experimental performance in terms of fuel economy and ABC events is well-balanced across experimental groups. A comparison of a rich set of trip-level and bus-type characteristics also reveals no differences across groups, indicating that the different groups have on average been exposed to very comparable driving conditions.[10] Drivers are on average 54 years old and have 20 years of experience at the company. Most drivers are male (89%). Trips have an average length of 31km and are typically driven in rural areas (84%).

The sample construction for the evaluation of the peer-comparison messages is discussed in detail in Romensen and Soetevent (2024). The final dataset consists of 842,845 observations at the driver-trip level. These trips are completed on of the three bus types operated by the company in the concession area: VDL, Intouro and IRIS. The VDL bus is used on most trips (75%) and runs, like the Intouro bus, on diesel. The IRIS bus runs on natural gas and is only used in one base location. To estimate the treatment effects, we will therefore restrict attention to trips completed by either a VDL- or Intouro bus. Furthermore, to align with the sample period used in our companion paper, we report the treatment effects based on data until 30 April, 2016. This is because some coaches indicated that they no longer kept track of their coaching sessions after April 2016, making it difficult for us to control for coaching sessions after this date. To rule out that uncontrolled for coaching effects influence our results, we conservatively restrict the sample period to the period for which we have complete coach logs available.[11]

4 Results

How does the dosage of the number and nature of peer-comparison feedback messages affect worker productivity? As our experimental design includes a control group that is not exposed to these targeted peer-comparison messages, we can adopt a difference-in-differences (DID) ap-

[9]For the ABC outcomes, Figure B.2 in the Online Appendix shows that half of all drivers manages to be in the top-25% at least once (panels a-c) and, depending on which dimension is considered, only 20 to 30 percent of all drivers always is in the bottom-50%. Online Appendix section B shows per treatment condition the number of messages send per month and the driving dimensions targeted.

[10]For each of the dimensions along which we stratified (base location, gender, years of service), $p \geq 0.99$.

[11]As an additional check, we also estimated the treatment effects based on the full sample period until 31 January, 2017. The main results do not significantly change if we consider this extended period. See, for example, Online Appendix section C.

proach that compares the driving behavior of control drivers with the short- and long-run performance of treated drivers in groups T2-T4. Specifically, we follow the approach in Romensen and Soetevent 2024 and estimate the following regression specification:

$$Y_{its} = \mu_i + \lambda_t + \gamma \mathrm{PF}_{it} + \sum_{j=2}^{4} \tau_j \mathrm{PF}_{it} \times \mathbf{1}(T_i = j) + X_{its} \cdot \theta + \kappa_b + \zeta_{bt} + \upsilon_{its}. \quad (1)$$

In this specification, Y_{its} refers to the outcome variable (fuel economy or the ABC dimensions) and is indexed by driver (i), time in days (t), and the bus trip (s). The variable PF_{it} captures the post-feedback period for a driver and equals one if driver i has received the first feedback report, zero otherwise. A no-report indicator is included in the specification as well to capture drivers operating after December 15, 2015, who were absent in the month on which the first report is based. The indicator $\mathbf{1}(T_i = j)$ indicates which experimental condition j a driver is assigned to. The τ-coefficients then capture the treatment effects of the targeted peer-comparison feedback messages. Control variables are included in the vector X_{its}. We control for travel distance, number of passengers and bus stops, rush hours, and non-scheduled rides. Importantly, since coaching takes place at the same time as the implementation of the feedback reports, we use the coach logs to pin down driver-specific coaching sessions and include post-coaching dummy variables as well. Our preferred specification also includes driver (μ_i), bus type (κ_b), day (λ_t), and bus type interacted with day (ζ_{bt}) fixed effects.

Table 3 shows for a number of specifications with different sets of controls and fixed effects the average treatment effects for fuel economy. The estimates show no significant effect of the peer-comparison messages in the text boxes. The point estimates across treatments are small in size, ranging from -0.086 to 0.069 liters/100km, and are individually and jointly insignificant. While fuel economy is an important outcome variable, the experimental feedback messages do not mention fuel economy but dissect a driver's relative performance into his/her performance on the disaggregate comfort dimensions acceleration, braking and cornering. The absence of a treatment effect for fuel economy need not imply that there is no effect at these 'lower' levels of driving behavior that are explicitly targeted by the intervention. However, the estimates in Table 4 show no consistent treatment effects for the ABC comfort dimensions either. The absence of an effect of peer-comparison feedback on conservation efforts among workers is consistent with findings in Blader et al. (2020). They note however that a focus on aggregate effects may mask temporal effects and improved performance among subgroups of drivers. In our case, estimates

of the effects of peer-comparison feedback for each month separately do not suggest the presence of such temporal effects. We find no indication that drivers respond differently to the first peer-comparison messages than to the ones received in later months, when they may lose attention.[12]

By design, only the subset of drivers who actually belong to the top-25% or bottom-50% receive a message. This makes it possible that we overlook some treatment effects among treated subgroups. The treatment estimates presented so far estimate the overall effect of being assigned a peer-comparison feedback treatment condition. This intention-to-treat (ITT) estimate is a conservative estimate of the average effect of actually receiving positive or negative messages.[13] In an exploratory analysis we consider per ABC outcome the subsample of drivers who rank in the bottom-50% or top-25% in a given month. We examine whether a driver's performance on that outcome improves in the next month after receiving corrective (when in the bottom-50%) or positive (when in the top-25%) peer-comparison feedback on that outcome this month. We compare, within subsample, the next month's rank of drivers in a treatment group (who received peer-comparison feedback in the current month) versus drivers in the control group (who did not receive feedback, despite being in the bottom or top bin). Table 5 summarizes the coefficient estimates. For acceleration and cornering, the effects remain insignificant. For braking, the estimates show positive effects of corrective feedback for all treatments, with improvements in ranking of 1.1 to 4.3 percentage points but none of these estimates is significant at the $p = 0.05$-level. For drivers in the top-25%, the coefficients are positive and sometimes significant, suggesting that a driver's ranking on a performance indicator deteriorates after having received positive feedback.

5 Conclusion

Existing studies have documented heterogeneous responses to the provision of relative performance feedback (Ashraf et al. 2014, Bandiera et al. 2013, Barankay 2012) and suggested that customizing feedback is a possible solution to encourage all workers to work hard to achieve a good rank (Kuhnen and Tymula 2012). However, how exactly relative performance feedback should be customized in the workplace to improve worker productivity remains an open question.

[12] Online Appendix section C shows the treatment effects per feedback round for the full sample period.

[13] For example, every month only about 70% of all drivers in treatments T2[1n0p] and T4[3n0p] actually receive messages in their text box. In treatment T3[1n1p], about all drivers (97%) have their text box filled with a message, but with variation in whether the box contains only corrective feedback (54%), only positive feedback (25%) or a combination of both (21%). See Table B.1 for detailed information on the composition of messages by treatment.

We partner with a large bus company to implement a field experiment on targeted peer-comparison feedback. We exploit innovations in on-board monitoring to introduce tailored feedback messages on specific productivity dimensions over which workers have direct control. At negligible marginal cost to the employer we target workers that can improve the most. Using detailed driver-trip level data on several productivity dimensions, we show that dosing corrective feedback or accompanying it with positive feedback is generally ineffective in improving driving behavior. In particular, receiving feedback on concrete outcomes on which they can improve does not seem to motivate poor-performing workers to improve, even when the feedback comes with praise for performance on other dimensions.

Our findings contribute to the literatures on relative performance feedback and incentive design for energy conservation efforts among workers. The adoption of digital monitoring systems in the workplace creates many new opportunities for the personalization of feedback. As our results suggest, however, tailoring and targeting feedback to individual productivity dimensions does not necessarily improve worker effort. Further research is needed to determine the most effective types of personalized feedback for improving productivity.

References

Abadie, Alberto, Susan Athey, Guido W. Imbens, and Jeffrey M. Wooldridge, "When Should You Adjust Standard Errors for Clustering?," *Quarterly Journal of Economics*, February 2023, *138* (1), 1–35.

Ashraf, Nava, Oriana Bandiera, and Scott S. Lee, "Awards Unbundled: Evidence from a Natural Field Experiment," *Journal of Economic Behavior and Organization*, April 2014, *100*, 44–63.

Azmat, Ghazala and Nagore Iriberri, "The Importance of Relative Performance Feedback Information: Evidence from a Natural Experiment Using High School Students," *Journal of Public Economics*, 2010, *94* (7-8), 435–452.

Bandiera, Oriana, Iwan Barankay, and Imran Rasul, "Team incentives: Evidence from a firm level experiment," *Journal of the European Economic Association*, October 2013, *11* (5), 1079–1114.

____ , Valentino Larcinese, and Imran Rasul, "Blissful ignorance? A natural experiment on the effect of feedback on students' performance," *Labour Economics*, June 2015, *34*, 13–25.

Barankay, Iwan, "Rank Incentives: Evidence from a Randomized Workplace Experiment," *Working paper*, 2012.

Benabou, Roland and Jean Tirole, "Self-Confidence and Personal Motivation," *The Quarterly Journal of Economics*, August 2002, *117* (3), 871–915.

Blader, Steven, Claudine Gartenberg, and Andrea Prat, "The Contingent Effect of Management Practices," *Review of Economic Studies*, March 2020, *87* (2), 721–749.

Blanes i Vidal, Jordi and Mareike Nossol, "Tournaments without Prizes: Evidence from Personnel Records," *Management Science*, 2011, *57* (10), 1721–1736.

Bloom, Nicholas and John van Reenen, "Measuring and explaining management practices across firms and countries," *Quarterly Journal of Economics*, November 2007, *122* (4), 1351–1408.

____ , Benn Eifert, Aprajit Mahajan, David McKenzie, and John Roberts, "Does Management Matter? Evidence From India," *Quarterly Journal of Economics*, 2013, *128* (1), 1–51.

Clark, Andrew E. and Andrew J. Oswald, "Satisfaction and comparison income," *Journal of Public Economics*, September 1996, *61* (3), 359–381.

Delfgaauw, Josse, Robert Dur, Joeri Sol, and Willem Verbeke, "Tournament Incentives in The Field: Gender Differences in The Workplace," *Journal of Labor Economics*, 2013, *32* (2), 305–326.

Dohmen, Thomas, Armin Falk, Klaus Fliessbach, Uwe Sunde, and Bernd Weber, "Relative versus absolute income, joy of winning, and gender: Brain imaging evidence," *Journal of Public Economics*, 2011, *95* (279-285).

Falk, Armin and Andrea Ichino, "Clean Evidence on Peer Effects," *Journal of Labor Economics*, January 2006, *24* (1), 39–57.

Fischer, Mira and Dirk Sliwka, "Confidence in knowledge or confidence in the ability to learn: An experiment on the causal effects of beliefs on motivation," *Games and Economic Behavior*, 2018, *111* (122-142).

____ and Valentin Wagner, "Do timing and reference frame of feedback influence high-stakes educational outcomes?," *Economics of Education Review*, June 2023, *94*, 102379.

Kandel, Eugene and Edward P. Lazear, "Peer Pressure and Partnerships," *Journal of Political Economy*, August 1992, *100* (4), 801–817.

Kuhnen, Camelia M. and Agnieszka Tymula, "Feedback, Self-Esteem, and Performance in Organizations," *Management Science*, 2012, *58* (1), 94–113.

Lazear, Edward P. and Sherwin Rosen, "Rank-Order Tournaments as Optimum Labor Contracts," *Journal of Political Economy*, 1981, *89* (5), 841–864.

London, Manual, *Job Feedback: Giving, Seeking and Using Feedback for Performance Improvement*, Lawrence Erlbaum Associates, Mahway, NJ., 2003.

Luttmer, Erzo F.P., "Neighbors as Negatives: Relative Earnings and Well-Being," *The Quarterly Journal of Economics*, August 2005, *120* (3), 963–1002.

Mas, Alexandre and Enrico Moretti, "Peers at Work," *American Economic Review*, March 2009, *99* (1), 112–145.

Rilke, Rainer Michael, Victor van Pelt, Sebastian Lehnen, and Christina Guenther, "Motivating Low Performers with Input-based Relative Performance Feedback: Evidence From a Field Experiment," Available at SSRN: `https://ssrn.com/abstract=3978948orhttp://dx.doi.org/10.2139/ssrn.3978948` December 13 2022.

Roels, Guillaume and Xuanming Su, "Optimal Design of Social Comparison Effects: Setting Reference Groups and Reference Points," *Management Science*, March 2014, *60* (3), 606–627.

Ryan, Richard M. and Edward L. Deci, "Self-Determination Theory and the Facilitation of Intrinsic Motivation, Social Development, and Well-Being," *American Psychologist*, 2000, *55* (1), 68–78.

Soetevent, Adriaan R. and Gert-Jan Romensen, "The Short and Long Term Effects of In-Person Performance Feedback: Evidence from a Large Bus Driver Coaching Program," University of Groningen working paper 2024.

Song, Hummy, Anita L. Tucker, Karen L. Murrell, and David R. Vinson, "Closing the Productivity Gap: Improving Worker Productivity Through Public Relative Performance Feedback and Validation of Best Practices," *Management Science*, June 2018, *64* (6), 2628–2649.

Staats, Bradley R., Hengchen Dai, David Hofmann, and Katherine L. Milkman, "Motivating Process Compliance Through Individual Electronic Monitoring: An Empirical Examination of Hand Hygiene in Healthcare," *Management Science*, May 2017, *63* (5), 1563–1585.

Syverson, Chad, "What determines productivity?," *Journal of Economic Literature*, 2011, *49* (2), 326–365.

Tran, Anh and Richard Zeckhauser, "Rank as an Inherent Incentive: Evidence from a Field Experiment," *Journal of Public Economics*, 2012, *96* (9-10), 645–650.

Tables

Table 1: Experimental Conditions

Conditions	General feedback	Max # positive message(s)	Max # negative message(s)	# drivers
T1 [0n0p]	Yes	0	0	103
T2 [1n0p]	Yes	0	1	102
T3 [1n1p]	Yes	1	1	102
T4 [3n0p]	Yes	0	3	102

Table 2: Descriptive Statistics of Experimental Conditions

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
	Full sample		**T1 (0n0p)**		**T2 (1n0p)**		**T3 (1n1p)**		**T4 (3n0p)**		Joint test: treatment	
	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.	effect = 0	
Pre-experimental performance												
Fuel economy	25.20	(2.37)	25.28	(2.21)	25.00	(2.47)	25.28	(2.50)	25.24	(2.31)	0.33	[0.80]
Acceleration	12.30	(6.43)	12.38	(6.65)	11.91	(6.59)	12.28	(6.27)	12.61	(6.27)	0.20	[0.90]
Braking	3.00	(4.39)	3.20	(4.63)	2.98	(4.60)	2.59	(3.66)	3.25	(4.63)	0.46	[0.71]
Cornering	2.88	(4.93)	2.99	(5.11)	2.85	(5.17)	2.62	(4.60)	3.06	(4.89)	0.16	[0.93]
Demographics												
Year of birth	1962	(8.28)	1962	(8.83)	1962	(8.62)	1962	(7.62)	1962	(8.12)	0.01	[1.00]
Year of employment	1996	(11.53)	1996	(11.65)	1996	(11.48)	1996	(11.32)	1996	(11.83)	0.04	[0.99]
% share of FTE≥0.9	76.28		75.73		74.51		79.41		75.49		0.26	[0.86]
% share of female drivers	10.51		10.68		9.80		10.78		10.78		0.02	[0.99]
Trip-specific variables												
Punctuality	-2.88	(0.80)	-2.80	(0.86)	-2.94	(0.76)	-2.80	(0.86)	-2.96	(0.72)	1.13	[0.33]
Distance traveled	31.36	(13.21)	31.32	(13.99)	31.89	(12.59)	31.14	(12.84)	31.10	(13.53)	0.08	[0.97]
Number of passengers	13.26	(3.71)	13.38	(3.91)	13.27	(3.69)	13.18	(3.60)	13.23	(3.69)	0.05	[0.98]
Number of bus stops	37.84	(8.79)	37.38	(9.09)	38.54	(8.40)	37.93	(8.76)	37.51	(8.99)	0.35	[0.79]
% share of rides:												
- Morning rush hours	19.45		19.34		20.53		18.64		19.31		0.46	[0.71]
- Evening rush hours	19.53		19.99		18.20		20.55		19.39		0.72	[0.54]
- Weekend	14.09		14.31		14.01		3.98		14.03		0.03	[0.99]
- Holidays	9.99		9.88		10.27		9.71		10.10		0.27	[0.85]
- Urban area	15.84		16.77		13.68		15.89		17.00		0.19	[0.91]
- School	0.8		0.7		0.8		0.8		0.7		0.21	[0.89]
% share of trips on bus types												
Bus type VDL	75.20		73.64		77.71		75.63		73.88		0.33	[0.81]
Bus type Intouro	9.65		10.21		9.34		9.30		9.75		0.10	[0.96]
Bus type IRIS	15.15		16.15		12.96		15.07		16.37		0.21	[0.89]
Base locations (# drivers)												
Location 1	12		3		3		3		3		0.00	[1.00]
Location 2	61		16		15		15		15		0.01	[1.00]
Location 3	30		7		8		8		7		0.05	[0.99]
Location 4	74		19		18		18		19		0.02	[1.00]
Location 5	150		37		38		38		37		0.02	[1.00]
Location 6	82		21		20		20		21		0.02	[1.00]
Number of drivers	409		103		102		102		102			

Notes: Unit of observation is the driver. Columns (11) and (12) show F-statistics and [p-values] from a balance test of whether the treatment coefficients are jointly equal to zero. Pairwise t-tests with the control group $T1$ do not reveal any statistically significant differences in means for any of the variables for any of the treatment groups at the $p = 0.10$ level. Data are from EOBRs in buses and centralized databases with driver and trip characteristics. Fuel economy in liters/100km. Performance on the comfort dimensions (Acceleration, Braking and Cornering) as the number of events/10km. The pre-experimental period is the period before receiving the first feedback report. Punctuality is the difference in minutes between actual and planned driving time. Distance traveled is measured in kilometers. Number of passengers based on check-ins with public transport cards. VDL and Intouro buses have diesel engines, the IRIS bus runs on natural gas. Morning and evening rush hours are from 7-10am and 4-7pm, respectively. Holiday rides take place during public and school holidays. School rides are along routes with schools and universities as final destinations. School rides are along routes with schools and universities as final destinations. Stars indicate a statistically significant difference in means with the control group.

Table 3: Targeted Peer-Comparison Feedback Effects on Fuel Economy

Dependent variable:	**Fuel Economy**			
	(1)	(2)	(3)	(4)
Post-feedback × T2 (1n/0p)	0.010	0.069	0.068	0.067
	(0.101)	(0.088)	(0.087)	(0.087)
Post-feedback × T3 (1n/1p)	-0.029	0.006	0.000	0.004
	(0.107)	(0.106)	(0.104)	(0.105)
Post-feedback × T4 (3n/0p)	-0.086	-0.051	-0.049	-0.048
	(0.095)	(0.095)	(0.094)	(0.094)
$\hat{\gamma}$ (PF)	-0.490***	-0.282***	0.002	-0.109
	(0.068)	(0.073)	(0.077)	(0.080)
Constant	27.558***	23.749***	26.690***	26.733***
	(0.080)	(0.057)	(0.131)	(0.126)
R^2	.101	.432	.439	.447
Number of trip-level observations	734242	734242	734242	734242
Controls	No	Yes	Yes	Yes
Driver fixed effects	No	Yes	Yes	Yes
Day fixed effects	No	No	Yes	Yes
Bus type × day fixed effects	No	No	No	Yes

Notes: Identification of the written feedback treatment effects on driving performance. The time period under consideration is from 01/01/2015 to 30/04/2016. Treatments vary in the number of positive and negative peer-comparison messages on the comfort driving dimensions (acceleration, braking, cornering). Messages are targeted in the sense that they are only provided if a driver performs relatively poor (bottom 50%) or good (top 25%) compared to a reference group of colleagues. Fuel economy is measured in liters/100km. Standard errors are clustered by driver which is the unit of randomization (Abadie et al. 2023). Controls include: travel distance, number of passengers and bus stops, and dummies for bus type, morning and evening rush hours, non-scheduled rides and having been coached. A no-report indicator is included to capture drivers operating after December 15, 2015, but who have not yet received their first report. ***(**,*) : statistically different from zero at the 1%-level (5%-level, 10%-level).

Table 4: Targeted Peer-Comparison Feedback Effects on ABC Comfort Dimensions

Dependent variable:	**Acceleration**		**Braking**		**Cornering**	
	(1)	(2)	(3)	(4)	(5)	(6)
Post-feedback × T2 (1n/0p)	-0.013	-0.121	0.042	0.027	0.027	0.053
	(0.241)	(0.222)	(0.055)	(0.043)	(0.076)	(0.073)
Post-feedback × T3 (1n/1p)	-0.110	-0.022	0.015	0.043	0.085	0.118*
	(0.240)	(0.223)	(0.055)	(0.046)	(0.069)	(0.065)
Post-feedback × T4 (3n/0p)	-0.435*	-0.421*	-0.005	-0.007	-0.003	0.045
	(0.247)	(0.230)	(0.055)	(0.044)	(0.075)	(0.067)
$\hat{\gamma}$ (PF)	-2.226***		-0.869***		-0.408***	
	(0.151)		(0.038)		(0.052)	
Constant	9.942***	8.774***	1.307***	1.440***	1.073***	1.350***
	(0.216)	(0.200)	(0.049)	(0.063)	(0.090)	(0.059)
R^2	.067	.515	.151	.425	.032	.387
Number of trip-level observations	189626	189626	189626	189626	189626	189626
Controls	No	Yes	No	Yes	No	Yes
Driver fixed effects	No	Yes	No	Yes	No	Yes
Day fixed effects	No	Yes	No	Yes	No	Yes
Bus type × day fixed effects	No	Yes	No	Yes	No	Yes

Notes: The ABC outcome variables are measured as the number of events per 10 kilometers. Further: see notes to Table 3.

Table 5: Rank Change in Month Following Receiving Actual Feedback on That Productivity Outcome

Rank in month of message		bottom-50%			top-25%
Treatment		T2[0p1n]	T3[1p1n]	T4[0p3n]	T3[1p1n]
Outcome:	Acceleration	0.009	0.012	0.008	0.028
	Braking	-0.043*	-0.027	-0.011	0.078**
	Cornering	-0.000	-0.010	-0.004	0.016

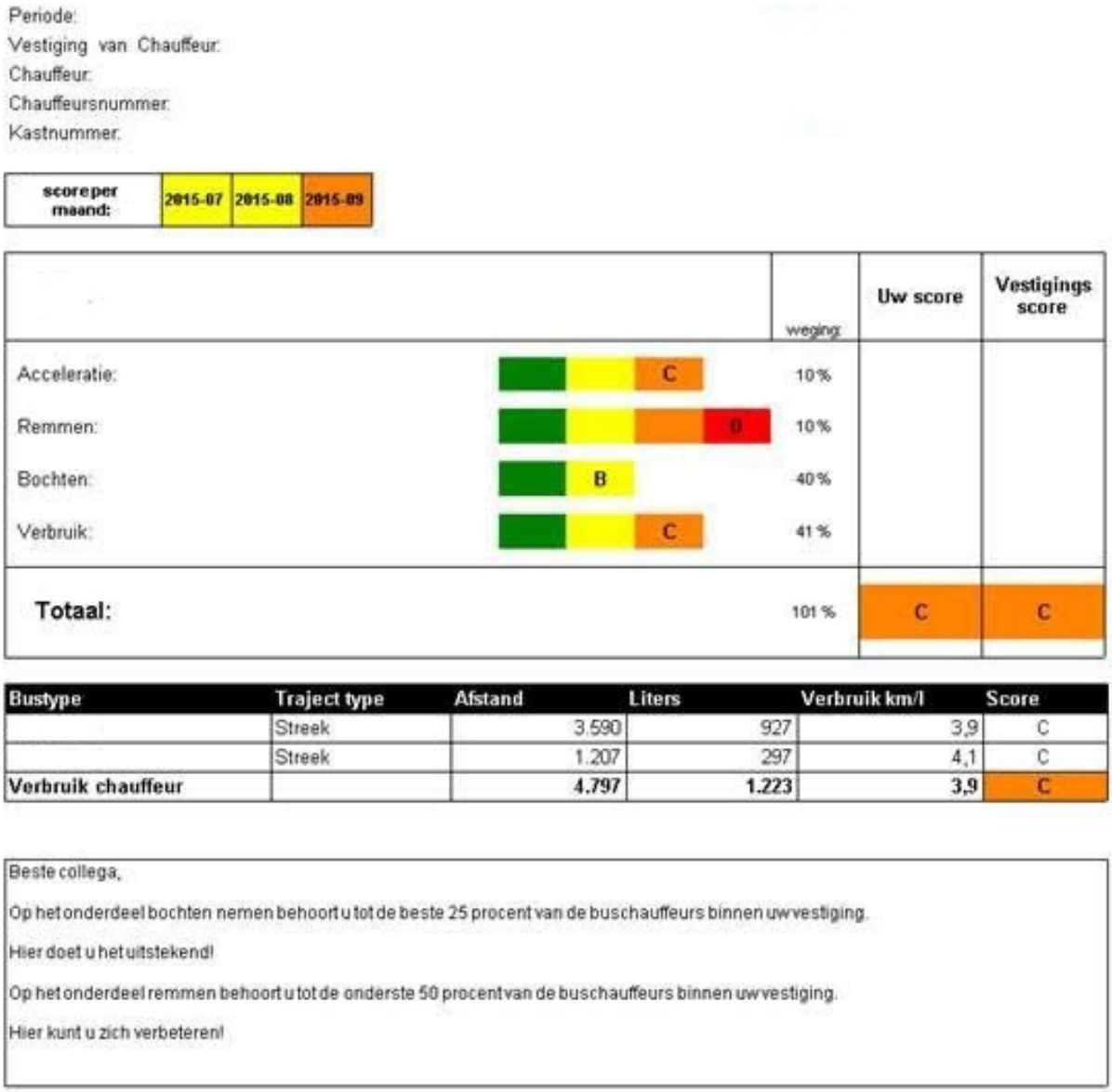
Notes: The best (worst) performing driver has rank 0 (1). A constant and treatment and month fixed effects are included in all regressions (full regression details in Appendix Table D.2). ***(**,*) : statistically different from zero at the 1%-level (5%-level, 10%-level). Standard errors are clustered by driver.

Online Appendix [Not for Publication]

A Sample feedback report

Figure A.1 reproduces a specimen of the feedback report drivers received once a month between December 2015 - November 2016.

Figure A.1: Sample Feedback Report



Periode:
Vestiging van Chauffeur:
Chauffeur:
Chauffeursnummer:
Kastnummer:

score per maand:	2015-07	2015-08	2015-09

	weging	Uw score	Vestigings score
Acceleratie: C	10 %		
Remmen: D	10 %		
Bochten: B	40 %		
Verbruik: C	41 %		
Totaal:	101 %	C	C

Bustype	Traject type	Afstand	Liters	Verbruik km/l	Score
	Streek	3.590	927	3,9	C
	Streek	1.207	297	4,1	C
Verbruik chauffeur		**4.797**	**1.223**	**3,9**	**C**

Beste collega,

Op het onderdeel bochten nemen behoort u tot de beste 25 procent van de buschauffeurs binnen uw vestiging.

Hier doet u het uitstekend!

Op het onderdeel remmen behoort u tot de onderste 50 procent van de buschauffeurs binnen uw vestiging.

Hier kunt u zich verbeteren!

Notes: Confidential information (related to the driver) has been removed.

B Driver exposure to targeted feedback

The tailored nature of the messages is illustrated in Table B.1. Panel A reports the percentage share of drivers receiving one of the possible message combinations in each treatment and feedback round (conditional on receiving a feedback report).[14] It highlights the flexible design of the treatments. Drivers in the control group, treatment 1 (0n/0p), receive no tailored feedback messages. Each treated driver is assigned an individualized message combination which points to behaviors that require attention. In treatment 2 (1n/0p), for example, about 70% of the drivers receive a negative message in a given feedback round, meaning that they perform poorly compared to peers on one of the three comfort dimensions. The remaining 30% performs well on all dimensions and is therefore not notified with a message. Panel B details the composition of the message combinations and shows that all ABC dimensions are well-represented.

How often a treated driver is in the top-25% or bottom-50% on a given driving dimension is shown in Figure B.2. The figure plots the number of feedback rounds a driver is in the bottom or top part of the reference group divided by the total number of feedback rounds in which the driver received a feedback report. This gives an indication how often a driver is eligible for targeted messages. For many drivers it varies per round whether they were in the target groups. On each dimension, we observe that there are drivers who were always or never in the bottom (top) part. On acceleration, 19% (16%) of the treated drivers were never (always) in the bottom 50%. For the top 25%, the corresponding figures are 42% (9%). Outcomes are similar for braking and cornering.

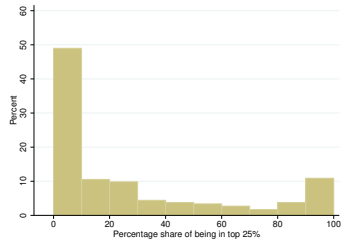
[14] No report is created when drivers were absent in the previous month (on which the report is based).

Table B.1: Incidence and Composition of Targeted Peer-Comparison Messages

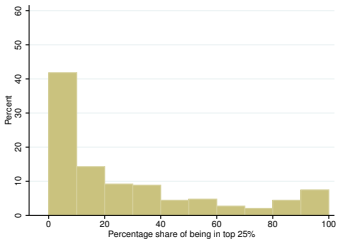
	T2 (1n/0p)		**T3 (1n/1p)**				**T4 (3n/0p)**			
Feedback round	0n/0p	1n/0p	0n/0p	1n/0p	0n/1p	1n/1p	0n/0p	1n/0p	2n/0p	3n/0p
	Panel A: incidence of messages									
December 2015	31%	69%	3%	53%	24%	20%	29%	26%	22%	23%
January 2016	27%	73%	7%	49%	22%	22%	29%	24%	17%	29%
February 2016	31%	69%	2%	49%	21%	28%	27%	27%	22%	24%
March 2016	30%	70%	3%	52%	23%	22%	32%	16%	24%	28%
April 2016	30%	70%	1%	53%	27%	19%	31%	18%	28%	23%
May 2016	29%	71%	2%	55%	27%	16%	28%	18%	34%	20%
June 2016	30%	70%	5%	51%	23%	22%	29%	19%	30%	22%
July 2016	26%	74%	3%	54%	24%	19%	25%	23%	26%	25%
August 2016	27%	73%	4%	47%	23%	25%	27%	24%	24%	24%
September 2016	32%	68%	2%	56%	26%	16%	28%	19%	34%	18%
October 2016	30%	70%	3%	52%	22%	23%	29%	22%	24%	25%
	Panel B: composition of messages (A%;B%;C%)									
December 2015		(31;46;23)		(40;23;38)	(33;39;28)	(80;33;87)		(30;35;35)	(65;59;76)	(100;100;100)
January 2016		(33;33;33)		(26;30;45)	(38;33;29)	(76;29;95)		(30;30;39)	(69;69;63)	(100;100;100)
February 2016		(20;41;39)		(34;43;23)	(55;15;30)	(67;59;74)		(44;20;36)	(52;81;67)	(100;100;100)
March 2016		(32;43;25)		(34;32;34)	(32;23;45)	(90;52;57)		(27;33;40)	(65;70;65)	(100;100;100)
April 2016		(33;38;30)		(27;37;35)	(42;23;35)	(56;72;72)		(35;24;41)	(63;74;63)	(100;100;100)
May 2016		(38;29;32)		(34;38;28)	(31;27;42)	(100;56;44)		(29;29;41)	(64;73;64)	(100;100;100)
June 2016		(40;35;25)		(31;37;33)	(45;27;27)	(67;62;71)		(28;28;44)	(68;71;61)	(100;100;100)
July 2016		(34;31;34)		(33;24;43)	(57;22;22)	(61;78;61)		(32;27;41)	(64;76;60)	(100;100;100)
August 2016		(38;30;33)		(29;36;36)	(32;32;36)	(88;58;54)		(43;30;26)	(61;61;78)	(100;100;100)
September 2016		(41;29;31)		(28;33;39)	(32;36;32)	(73;80;47)		(28;44;28)	(69;56;75)	(100;100;100)
October 2016		(47;25;27)		(39;27;35)	(38;24;38)	(50;86;64)		(50;25;25)	(55;68;77)	(100;100;100)

Notes: Panel A reports the percentage share of drivers receiving one of the possible message combinations in each treatment (conditional on receiving a feedback report). Drivers do not receive a report when they were absent in the previous month (on which the report is based). Panel B shows the composition of these combinations in terms of the ABC comfort dimensions. Negative (positive) messages are provided if a driver belongs to the bottom 50% (top 25%) of a reference group of peers on one of the comfort dimensions. The reference group consists of drivers who share the same base location and treatment status. Drivers in the control group, T1 (0n/0p), receive no peer-comparison messages. A printed version of the feedback report (with the messages integrated) is created around the 15th day of each feedback round and delivered via the team manager or pigeonhole.

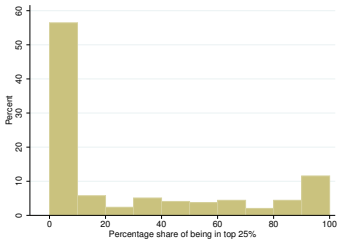
Figure B.2: Share of Feedback Rounds in Top 25% or Bottom 50% for Treated Drivers



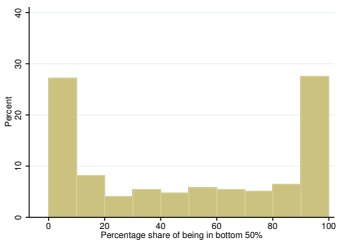
(a) Top 25%: acceleration



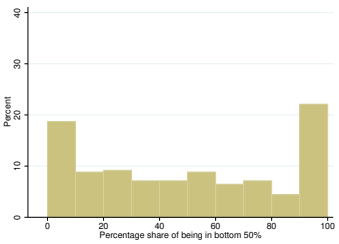
(b) Top 25%: braking



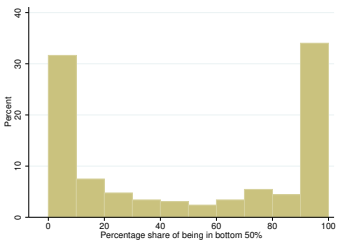
(c) Top 25%: cornering



(d) Bottom 50%: acceleration



(e) Bottom 50%: braking

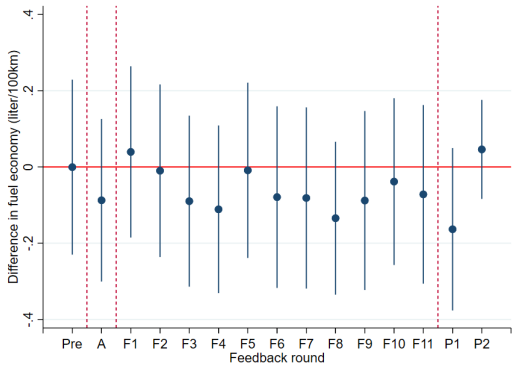


(f) Bottom 50%: cornering

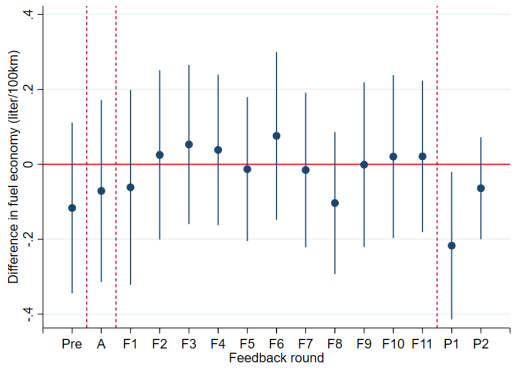
Notes: The figures show for each ABC driving dimension the distribution of feedback round shares in which treated drivers were in the top 25% or bottom 50% of the peer reference group. The shares are calculated as the number of feedback rounds a driver was in the bottom or top part of the reference group divided by the total number of feedback rounds in which a feedback report was constructed for the driver. It indicates how often a driver was eligible for a targeted peer-comparison message on a given driving dimension (the received message combination depends on the treatment condition). The reference group consists of drivers who share the same base location and treatment status.

C Temporal treatment effects

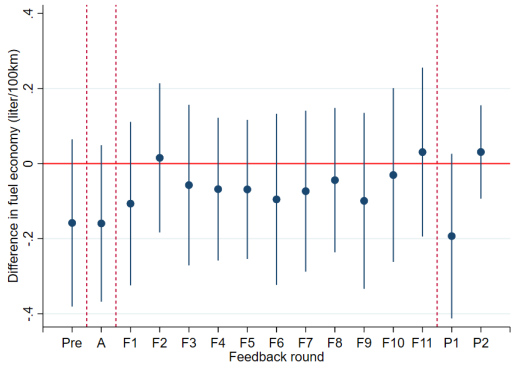
Figure C.3: Temporal Treatment Effects of Targeted Peer-Comparison Feedback Messages



(a) T2: 1n/0p



(b) T3: 1n1p



(c) T4: 3n0p

Notes: Treatment effects of targeted peer-comparison feedback messages per feedback round. The time period under consideration is from 01/01/2015 to 31/01/2017. Treatments vary in the number of positive and negative peer-comparison messages on the comfort driving dimensions (acceleration, braking, cornering). Messages are targeted in the sense that they are only provided if a driver performs relatively poor (bottom 50%) or good (top 25%) compared to a reference group of colleagues.

D Change in ranking following relative performance feedback

Table D.2: Rank change in month $(m+1)$, following peer-comparison feedback on that indicator in month m

Dep. var	Rank month $m+1$			
Treatment:	T2[0p1n]	T3[1p1n]	T4[0p3n]	T3[1p1n]
Sample (rank month m):	bottom-50%			top-25%
	Acceleration			
Message	0.009	0.012	0.008	0.028
	(0.020)	(0.018)	(0.017)	(0.021)
Rank month m	0.663***	0.645***	0.719***	1.188***
	(0.062)	(0.060)	(0.054)	(0.156)
obs.	698	722	921	355
R^2	0.211	0.201	0.235	0.243
	Braking			
Message	-0.043*	-0.027	-0.011	0.078**
	(0.023)	(0.022)	(0.019)	(0.035)
Rank month m	0.550***	0.550***	0.567***	0.624***
	(0.071)	(0.070)	(0.069)	(0.181)
obs.	711	710	934	314
R^2	0.141	0.144	0.138	0.115
	Cornering			
Message	0.000	-0.010	-0.004	0.016
	(0.011)	(0.011)	(0.009)	(0.016)
Rank month m	0.871***	0.873***	0.834***	0.838***
	(0.036)	(0.037)	(0.037)	(0.104)
obs.	683	727	944	336
R^2	0.532	0.513	0.463	0.120

Notes: The best (worst) performing driver has rank 0 (1). A constant and month fixed effects are included in all regressions.
***(**,*) : statistically different from zero at the 1%-level (5%-level, 10%-level). Standard errors in parentheses. Standard errors are clustered by driver.