university of groningen
faculty of economics and business

2020009-EEF

Improving Worker Productivity Through Tailored Performance Feedback: Field Experimental Evidence from Bus Drivers

April 2020

Gert-Jan Romensen
Adriaan R. Soetevent

university of groningen

faculty of economics and business

SOM is the research institute of the Faculty of Economics & Business at the University of Groningen. SOM has six programmes:

- Economics, Econometrics and Finance
- Global Economics & Management
- Innovation & Organization
- Marketing
- Operations Management & Operations Research
- Organizational Behaviour

Research Institute SOM
Faculty of Economics & Business
University of Groningen

Visiting address:
Nettelbosje 2
9747 AE Groningen
The Netherlands

Postal address:
P.O. Box 800
9700 AV Groningen
The Netherlands

T +31 50 363 9090/7068/3815

www.rug.nl/feb/research

Improving Worker Productivity Through Tailored Performance Feedback: Field Experimental Evidence from Bus Drivers

Gert-Jan Romensen
University of Groningen, Faculty of Economics and Business, Department of Economics, Econometrics and Finance

Adriaan R. Soetevent
University of Groningen, Faculty of Economics and Business, Department of Economics, Econometrics and Finance
a.r.soetevent@rug.nl

Improving Worker Productivity Through Tailored Performance Feedback: Field Experimental Evidence from Bus Drivers

Gert-Jan Romensen*
University of Groningen

Adriaan R. Soetevent†
University of Groningen

April 20, 2020

Abstract

How should performance feedback be tailored to improve worker productivity? In a natural field experiment with bus drivers, we test the potential of two forms of individual feedback: written peer-comparison feedback and in-person coaching.

We find that the announcement of the written feedback program has a substantial and significant effect on fuel economy and outcomes pertaining to passenger comfort; targeted peer-comparison feedback is generally ineffective; in-person coaching generates significant improvements on all dimensions for drivers in the bottom half of the performance distribution for about eight weeks; in-person coaching reduces the impact of written peer-comparison feedback but not vice versa.

JEL classification: D23, J24, M53, Q55.
Keywords: labor productivity, feedback, peer comparisons, field experiment.

*Corresponding author. 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, ESA Los Angeles 2019, IAEE 2018, 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 and the unique identifying number is: "AEARCTR-0005391".

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

1 Introduction

Giving effective performance feedback is critical in maintaining and enhancing worker productivity, especially in work environments that hinder the use of pay-for-performance schemes (Blader, Gartenberg and Prat 2020, Gosnell, List and Metcalfe 2020). The adoption of digital monitoring technologies at the work floor has made detailed individual-level data on disaggregated productivity measures available and hence greatly expanded managers' scope for giving workers tailored performance feedback (Staats, Dai, Hofmann and Milkman 2017). This increases the need to answer two important yet unsettled questions concerning optimal feedback provision. First, is feedback more effectively delivered in person or via automatically generated individual-specific feedback reports? The combination of finer data granularity and digital storage makes the latter feasible at low marginal cost. Second, which dimensions of worker productivity should the feedback target? The additional detail on the constituent parts of worker productivity gives managers more choice in selecting feedback intensity and in combining positive with negative feedback. Should they provide feedback on all dimensions simultaneously to prevent drivers from underperforming in non-reported dimensions (Hölmstrom and Milgrom 1991, Baker 1992) or should they instead limit feedback to prevent information overload (Simon 1973, Hitt and Brynjolfsson 1997, Edmunds and Morris 2000)?[1]

This paper aims to contribute to answering these questions. We run a field experiment at a large public transport company that is in the process of installing electronic on-board recorders (EOBRs) in its entire bus fleet. EOBRs enable the high-frequency measurement of a range of productivity outcomes, such as fuel efficiency and the number of Acceleration, Braking and Cornering events, the so-called ABC comfort dimensions.[2] Digital monitoring technologies such as EOBRs offer great potential in improving the quantitative evaluation of the effectiveness of different forms of performance feedback. Yet, this potential is thus far largely untapped. Feedback eligibility and feedback intensity are likely to correlate

[1]Recent studies that examine how the adoption of electronic monitoring technologies by firms impact worker productivity include Pierce, Snow and McAfee (2015) and Kelley, Lane and Schönholzer (2018).

[2]More generally, innovations in the transport sector related to on-board monitoring open up novel opportunities to measure worker productivity. See Baker and Hubbard (2003) and Hubbard (2003) for early work incorporating this technology. They study how the adoption of on-board computers has influenced the decision of truckers to integrate or outsource trucking services.

with workers' (relative) productivity outcomes. This sample selection biases estimates of feedback effectiveness that are based on comparisons of worker productivity just before and right after the worker has received feedback.

To avoid such bias, we combine detailed EOBR data from a sample of 409 bus drivers with random treatment variation in feedback format and feedback intensity. This creates a unique opportunity to quantitatively evaluate the effectiveness of different forms of performance feedback, allowing us to present estimates on the causal impacts of varying feedback intensity and feedback channel (written or in person) on worker productivity.

Following the launch of the company's EcoManager campaign to promote efficient and comfortable driving, all drivers receive a monthly written feedback report on their driving performance in the preceding month. This part of the campaign is not subjected to experimental variation: the launch date and timing of the monthly feedback are the same for all drivers. To this general report, we add a text box in which we experimentally vary the number of ABC dimensions on which drivers receive information on their relative ranking. This text box is empty for drivers in the control group. Drivers in the first treatment condition receive information on their poor relative performance (if any) on only one of the ABC dimensions, even when performance is relatively poor on multiple dimensions. That is, we deliberately withhold some rankings to allow drivers to focus their effort. The second treatment condition is similar, except that negative feedback is supplemented with positive feedback in case a driver who performs poorly on some dimensions scores well on others. This allows us to assess the value of providing a mix of corrective and positive feedback. In the final condition, all relative positions on driving behaviors are communicated whenever the driver performs poorly compared to a reference group of peers. Together, these interventions enable us to explore the potential of on-board monitoring technologies in customizing written relative performance feedback such that it enhances worker motivation.

In addition to the written peer-comparison feedback, we evaluate the effects of a parallel in-person coaching program with a quasi-experimental design. In this program, designated experienced drivers engage in coaching their colleagues by riding along with a bus driver for a portion of the driver's shift. At the end of the ride, the coach evaluates the trip in detail and gives tailored tips for improvement. Due to the hop-on hop-off approach to coaching and regulations that disallow coaches access to the driver's performance, the timing of the coaching sessions can be considered the outcome of quasi-random assignment: coaches select the drivers they will coach on a given day in a way that is unrelated to a driver's past performance. Our empirical evidence corroborates this. The (quasi-)random assignment of the different feedback designs thus avoids the aforementioned selection problems. We follow drivers for two years in order to establish a long baseline and experimental period. This enables us to measure both the immediate and delayed response to the feedback programs. We evaluate the two feedback formats using over 500,000 trip-level observations.

Our main findings are as follows. First, the launch of the general EcoManager campaign reduces fuel consumption by 0.4 liters/100km (0.40 standard deviation, SD). Distributing the feedback reports generates a further 0.1 SD reduction. For the peer-comparison feedback, we find precisely estimated zero effects. Varying the number and nature of peer-comparison feedback messages has no additional impact on worker productivity.

Second, we observe strong and immediate effects of coaching. On the day of coaching the fuel need reduces by 0.6 liters/100km (0.58 SD, p <0.001) and the number of acceleration events by 1.1 events/10km (0.50 SD, $p < 0.001$). For braking and cornering behavior, these effects are less pronounced and not (braking) or less (cornering) significant. The improvements due to coaching tend to persist with a smaller magnitude in the ensuing weeks but fade out after about seven to nine weeks. Zooming in, we find the impact of coaching on performance confined to drivers in the bottom half of the performance distribution.

Third, we find a nonreciprocal relation between in-person coaching and written peer-comparison feedback: prior exposure to peer-comparison messages does not change the effectiveness of in-person coaching for any of the productivity measures. Peer-comparison feedback, however, is only effective in the group of drivers that did not yet receive in-person coaching. One possible explanation is that once drivers have met a coach who gave them detailed feedback on what they do right and wrong on a trip, they become insensitive to subsequent written messages about their relative performance.

Fourth, in the group of non-coached drivers, those in the treatment with the maximum number of negative messages and no positive comments show the largest improvement in productivity outcomes. In other words, limiting negative feedback or mixing negative with positive feedback does not seem to have any beneficial effect. This shows that it is important to pay attention to interactions between the different elements of job design.

This paper proceeds as follows. Section 2 reviews the related literature. Section 3 describes the field setting of the study. Section 4 elaborates on the research design, provides further details on both feedback programs and presents the data. The empirical analysis of both programs follows in Section 5. Section 6 discusses the results and concludes.

2 Related Literature

A large literature shows that management practices matter for worker productivity (Bloom and van Reenen 2007, Bloom, Eifert, Mahajan, McKenzie and Roberts 2013, Syverson 2011). Despite a considerable body of empirical work, the question how relative performance feedback 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, Tucker, Murrell and Vinson 2018), sales growth (Delfgaauw, Dur, Sol and Verbeke 2013) and (high school) student performance (Tran and Zeckhauser 2012, Azmat and Iriberri 2010). Other studies, however, report decreased performance following the provision of rank information (Ashraf, Bandiera and Lee 2014, Bandiera, Barankay and Rasul 2013) and improved performance when they are abolished (Barankay 2012). People may exhibit rank incentives (Barankay 2012, Tran and Zeckhauser 2012) when relative performance information affects self-image (Benabou and Tirole 2006) and status (Moldovanu, Sela and Shi 2007). These rank incentives can lead to demotivation at the bottom of the performance distribution, which reduces the average effects of feedback programs that rely on social comparisons (Ashraf et al. 2014). Kuhnen and Tymula (2012) suggest that it may be promising to customize relative performance feedback by tailoring the content or by targeting subsets of workers. Blader, Gartenberg and Prat (2020) for example find that the provision of relative performance information in plants with(out) a teamwork culture leads to decreased (improved) truck driver performance.

What may account for some of the heterogeneity in results is that rankings are typically reported on final outcomes rather than on the intermediate steps leading to these outcomes. In this form, the message may be demotivating because it gives little guidance on where to improve and signals that improvement requires one big step rather than several small and clear steps. Feedback provision on disaggregated productivity measures can provide much more guidance on where to improve, making it easier for workers to change their behavior. It may empower poor performers by increasing the feeling of control, raising awareness of behaviors that require attention, and by offering suggestions for specific actions that workers can take. The feeling of being in control is a key source of human motivation (Ryan and Deci 2000).

Our research design does exactly that. One possible concern with disaggregated relative performance feedback, however, is that it may aggravate the adverse effects of feedback provision. That is, it may make poor performance even more salient to workers at the bottom of the distribution. When information directly enters the utility function (Golman, Hagmann and Loewenstein 2017), informing workers about poor performance on multiple dimensions may decrease motivation.[3] Also, the increased level of detail in the written feedback may trigger adverse effects similar to those caused by feedback overload. Increasing the feedback frequency can lead to more mistakes (Eriksson, Poulsen and Villeval 2009) and reduced task effort due to overwhelmed cognitive resources (Lam, DeRue, Karam and Hollenbeck 2011). This poses a challenge, as poor performers have the biggest room for improvement and are thus precisely the group that one wishes to target with detailed feedback.

Treatment effect heterogeneity may also show in the drivers' response to in-person coaching. A prevalent finding in the literature on peer effects in educational outcomes (Sacerdote 2011) is that high-ability students benefit most from the presence of high-ability peers (Fruehwirth, 2013, Hoxby and Weingarth, 2005, Lavy, Paserman, and Schlosser,

[3]Dohmen et al. (2011), for example, show that reward-related brain areas negatively correlate with lower relative incomes.

2011, Lavy, Silva and Weinhardt, 2012) although some studies (Burke and Sass 2013) find that students with the lowest past performance gain most from exposure to higher-achieving peers.[4] Drivers in our design are coached by experienced colleagues assigned the role of coach. Hence, a coaching session explicitly exposes a driver to a high-achieving peer. While recognizing the differences between a school environment and the work environment that we study – both in the nature of the interactions and the outcomes of interest – the cited studies suggest that the effect of in-person coaching may depend on a driver's own past performance. Our study checks whether this result on peer effects carries over to non-educational contexts. A related study is Sandvik, Saouma, Seegert and Stanton (2020) who run a field experiment among salespeople. They similarly find that exposure to a high-achieving peer generates productivity gains but in their setting, the gains persist even after twenty weeks.

Next to contributing to the empirical literature on optimal feedback design in operations management, our findings also address the broader societal challenge of how to combat unsustainable energy consumption practices. While there has been much progress in our understanding of non-financial incentives in residential energy consumption, research on how these insights generalize to firms is scant (Gerarden, Newell and Stavins 2017, Gosnell et al. 2020, Nilekani 2018).[5] Our work aims to partly fill this gap and should be viewed as part of the emerging literature that looks at the workplace for evidence on the effect of non-financial incentives on conservation efforts (Gosnell et al. 2020). Given that firms increasingly record and store data on multiple dimensions of worker-level productivity, tailoring feedback by decomposing consumption into its underlying sources seems

[4]Booij, Leuven and Oosterbeek (2017) find that low-ability students benefit from having low-ability peers but that high-ability students are unaffected by their peer group composition.

[5]Existing studies on non-financial incentive schemes in the residential sector stress the importance of feedback and social approval in increasing welfare (Allcott and Mullainathan 2010). For example, incorporating social comparisons in feedback reports reduces household consumption of energy (Allcott 2011, Ayres, Rasemand and Shih 2013) and water (Ferraro and Price 2013), with long-run effectiveness depending on whether households alter their capital stock of habits or physical technologies (Allcott and Rogers 2014). Recent research, however, also notes that social comparisons can trigger asymmetric effects (Holladay, LaRiviere, Novgorodsky and Price 2016) and may interact with other non-financial incentives when stimulating green behavior (Hahn, Metcalfe, Novgorodsky and Price 2016). This has reinforced the need for detailed evaluations of non-financial incentives pertaining to energy efficiency and also raises the question how these findings generalize to workers. Allcott and Kessler (2019) emphasize the importance of incorporating the (moral and emotional) costs incurred by nudge recipients in assessing the welfare effects of social comparisons.

a viable and promising approach to creating novel data-driven designs of conservation incentives (Brynjolfsson and McElheran 2016). The setting of a transport company is apt as the transport sector takes a heavy toll on the environment, accounting for one-fifth of global primary energy use and one-quarter of energy-related carbon dioxide (CO_2) emissions (IEA 2012). Indeed, the International Council on Clean Transportation hails fuel-efficient driving as low-hanging fruit to improve conservation levels (ICCT 2013).[6] However, picking this fruit can be challenging when drivers have no financial stake.

3 Field Setting

3.1 Industry

Our field partner is Arriva, a European-wide passenger transport company operating various transport modes in public transport. Bus transport is the firm's largest business unit.[7] In the Netherlands, bus concessions are granted to companies by means of a tendering procedure.[8] Winning a tender gives companies the exclusive rights to operate in a designated area for a number of years. To stimulate firms to engage in environmentally friendly behavior and to improve the living conditions of its citizens, local governments let environmental objectives feature prominently in the requirements tendering parties need to meet.[9] This has geared public transport companies toward the use of environmentally friendly technologies.[10] In the long run, this trend may drive bus companies to buy vehicles with a hybrid or electric fuel technology. On a shorter time horizon, the installment of electronic on-board recorders (EOBRs) helps the companies to meticulously measure performance on several dimensions of driving behavior. For example, the version used by Arriva records trip-level performance on fuel consumption and comfort dimensions such

[6]Barkenbus (2010) has sketched the potential of multidimensional eco-driving campaigns and feedback mechanisms for personal transportation. We instead examine the extent to which this potential can be realized in public transportation.

[7]At the time of the study, Arriva Group is part of Deutsche Bahn, employs over 60,000 people and annually delivers more than 2.2 billion passenger journeys in 14 European countries.

[8]See the Passenger Transport Act 2000.

[9]Interested companies are commonly requested to submit a sustainability plan in which they indicate how they decrease the ecological footprint of public transport in the concession area.

[10]The Dutch Ministry of Infrastructure mentions public transport as a "trend setter" in the area of sustainable technologies (MIVW, 2010, p. 87).

as acceleration, braking and cornering (ABC). Each driver logs into the system with a unique personnel number to match the performance records and trip-related background variables. This enables precise monitoring and provides managers and researchers with a wealth of high-frequency data on worker productivity and conservation efforts.

The system works as follows for the comfort dimensions. Based on test rides under different circumstances, threshold performance levels are formulated by the company for every dimension. Technically, the thresholds relate to minimum G-force measurements by a three-axis accelerometer in the bus. During each trip, the EOBR records an 'event' whenever an action by the driver is in excess of these thresholds. The performance measure of the ABC dimensions is the number of events per 10km, with fewer events indicating better driving behavior. The outcome data can subsequently be linked with centralized databases containing information on a host of driver and trip characteristics. This allows us to get a detailed picture of driver performance over time under various on-the-road conditions.

3.2 Research Setting

As part of its EcoManager campaign, Arriva Netherlands installed new EOBRs in its entire fleet in the time period 2015-2017. The EOBR data will be used as input to monthly feedback reports that will be distributed among the drivers. In addition, a new coaching program is introduced in which drivers receive real-time feedback and advice from an experienced colleague during on-the-road sessions. The new technology and the feedback programs are phased in over time in the concession areas.

We join the implementation process in the first concession area, comprising about two-thirds of a province in the Netherlands and serving about 5.16 million travelers in a year.[11] The majority of drivers in this area are tenured employees, while a small number (about 14%) operates on a temporary contract. Most of the drivers are experienced and have a long career of driving buses or other vehicles. They are typically not involved in other tasks within the organization. Opportunities for promotion are limited and the

[11] Based on the official number of electronic check-ins with the public transport card in 2015.

work council is against using financial incentives to reward good performances.[12] In the past, drivers received no personal feedback.

Each driver belongs to one of the six base locations (usually a municipality) in the area and operates on routes that are stipulated by the concession. For five locations, virtually all routes are between cities and in rural areas. Routes are based on timetables and do not vary much over time. One location (the largest one) has a mixture of urban and rural routes. Urban trips are mostly operated by a special bus type that runs on natural gas. Within a location, drivers' weekly shifts rotate. This implies that the worker faces week-to-week variation in his or her assignment to trips and the schedule repeats after about 14 weeks. This way of scheduling ensures that drivers are familiar with their routes and drive each route under different on-the-road circumstances. The schedules provide ample within-location variation in the type of trips, such that all drivers face a more or less similar mixture of relatively easy and difficult trips. Because of the rotation of shifts multiple drivers are assigned a given route. Together this variation allows us to include a rich set of fixed effects in our empirical analysis.

3.3 Scope for Improvement

Before discussing the research design, we wish to get an idea of the potential scope for improvement by considering the factors that influence driver performance on fuel economy and the ABC dimensions. What part of performance can be influenced by the driver and what part is caused by external factors such as weather and traffic conditions? For fuel economy, we observe sizable between-driver variation in performance. To drive 100km, the average driver uses 24.91 liters of fuel, with a standard deviation of $\sigma = 2.30$.[13] Table 1 shows that part of this variation can be attributed to differences in driving conditions.

The first column shows that the bus type accounts for 27.9 percent of the between-

[12] Within firms, the design of conservation incentives is often dictated by institutional constraints that hinder the use of pay-for-performance schemes. See e.g. Freeman (1981), who finds that within-establishment dispersion of wages is narrower in unionized establishments. He attributes this in large part to unions' wage practices, such as the adoption of uniform wages (rather than merit-based pay).

[13] 25 liters/100km $\sim$ 10.6 gallon/100miles. Throughout the text, we will state (changes in) fuel economy in l/km instead of km/l because of the miles-per-gallon (MPG) illusion (Larrick and Soll 2008). Figure A3 shows the entire distribution of driver fixed effects for the outcome variable fuel economy.

trip variation in fuel economy, with the Intouro and longer buses having a sizable and significantly worse fuel economy. The impact of weather conditions (column (2)) seems limited. Fuel economy is – as one expects – negatively correlated with the number of stops per kilometer, the number of passengers, evening rush hours and the bus running late. These variables seem to capture most of the day-to-day variation in fuel economy, as adding day fixed effects only slightly improves the R^2. Structural differences in driver performance explain an additional eight percentage points of variation in trip-level fuel economy (column (5)). When we control for the rich set of trip characteristics as in column (5) of Table 1, the variation in performance between drivers as measured by the residual standard deviation is $\sigma_r = 1.03$.[14] Hence, the potential for improvement is economically significant: A policy able to move a driver's average fuel economy from the 90th percentile to the 10th percentile reduces this driver's fuel bill by 2.46 liters/100km or about 10%.

We use the residual variation σ_r to compute the coefficient of variation $c_v = \sigma_r/\mu$ as a standardized measure of dispersion to compare the relative scope for improvements in fuel economy and in the ABC dimensions. For fuel economy, this coefficient equals 0.04 (= 1.03/24.91). The numbers for the ABC dimensions are shown in Table 2. The coefficients of variation show that in relative terms, between-driver dispersion is larger for the ABC dimensions than for fuel economy. However, for braking and cornering the average number of events per 10km is relatively close to the absolute lower bound of zero, thereby limiting the upward potential for a large fraction of drivers.

Of course, the different outcomes are related: more acceleration events for instance increase fuel consumption. Table 2 shows the residual correlation between the fuel economy and comfort dimensions after controlling for the same set of trip-level characteristics as in column (4) of Table 1. Fuel economy is correlated with acceleration and, to a lesser extent, with cornering. This supports the focus on the ABC dimensions in our peer-comparison treatments. Next to being worker productivity measures in their own right, improvements in either of them also contribute to fuel economy.

[14] Appendix section D.1 provides detail on the estimation of σ_r and section D.2 contains the corresponding tables for the ABC dimensions.

4 Experimental Design

4.1 Time Path

Figure 1 depicts the timeline of the study. First, we use the old on-board system to establish a long baseline of fuel consumption, starting in January 2015. At this stage, drivers are not informed about the upcoming feedback, nor that they are being monitored. The new EOBR system enables the collection of comfort dimensions baseline data in the months September and October 2015. The company sent promotion material about the EcoManager-project to the different locations on October 5, 2015. The project was officially launched with a kickoff event at November 9, 2015. At this date, the LED-array in the buses is also switched on, providing drivers with some instant feedback.[15] At the event, all drivers were informed about the digital monitoring and the introduction of monthly individualized feedback reports starting in December 2015.

Peer-comparison feedback The second period (Nov. 9-Dec. 15, 2015) is used to disentangle effects of the announcement and LED activation from the feedback effect. In the third period (Dec. 15, 2015-Nov. 15, 2016) drivers receive their monthly feedback reports with peer-comparison feedback. Finally, the post-experimental period (Nov. 15, 2016 - Jan. 31, 2017) starts with a one-time notification to the drivers that the peer-comparison messages are no longer included in the reports.[16]

Previous research has shown that workers adjust their effort in response to a feedback announcement, even though they have not yet learned any new information from the first feedback round (Blanes i Vidal and Nossol 2011). The company's decision to separate these events is a convenient feature of our research setting. Drivers were informed during the announcement period that the feedback will not be used in formal evaluations. This

[15]The LED-array contains eight LEDs: three green, two amber and three red. The green LEDs illuminate when the driver is in the 'sweet spot zone', determined by the (vehicle dependent) rotations per minute of the engine. The LEDs indicate the occurrence of an ABC event by flashing three times one second. As these events can only be timed when an action by a driver exceeds the threshold, any LED-array indication happens ex post.

[16]The precise text of this message 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". This message was part of the report that was distributed in November 2016 to all drivers that were part of the treatment conditions with peer-comparison feedback (all drivers except those in the control condition).

may rule out career concerns as an alternative explanation, but note that it runs counter to the firm's objectives to follow through on this claim (Hölmstrom 1979).

Apart from the feedback programs under consideration, no other incentives were used by the company to promote conservation efforts among workers. In the spirit of Barankay (2012), the one-time notification message is included at the end of the experiment in order to examine the effect of a withdrawal of peer-comparison messages.

In-person coaching The face-to-face coaching program that runs in parallel starts around the kickoff event in November 2015. Most drivers receive their first coaching in the weeks following the kickoff event.[17] During this period, the company reserved extra time for the coaches to ride along with drivers and to answer questions related to the upcoming feedback. Coaching intensity gradually decreases until it levels off after the first feedback report in mid December 2015. In a few cases, drivers participated in additional coaching sessions (55 drivers, 18% of all coached drivers). We control for these additional sessions in our analyses. We have complete coach logs for the period till April 30, 2016. Some coaches indicated that they no longer provided or kept track of coaching after April 2016. For this reason, we restrict attention to the period till April 30, 2016, in our evaluation of the coaching program. Thirty-two drivers (10% of all coached drivers) received coaching prior to the feedback announcement.

4.2 Peer-Comparison Treatments

Fueled by the conviction that the biggest gains in fuel-efficient and comfortable driving can be made when behaviors with the largest room for improvement are targeted, the company wants the peer-comparison feedback messages to emphasize the dimensions on which the driver can improve. The treatment variation in peer-comparison messages is integrated into the monthly feedback report received by all drivers.[18]

Drivers are randomly assigned an experimental condition, stratified along the dimensions of base location, gender, and years of service at the company. We construct reference

[17] See Figure A4 of the online Appendix.
[18] A sample feedback report is provided in Figure A2 of the online Appendix.

groups in which driver performance on each comfort dimension is compared to colleagues with the same base location and treatment status.[19] This creates a natural and homogeneous comparison group for drivers 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.[20] 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.

Treatment T1 [0n0p] is the control condition with no peer-comparison messages. In treatment T2 [1n0p], one negative message is provided if drivers underperform on a particular 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 made aware of 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 run the risk of receiving corrective feedback on all comfort dimensions. Using T3 [1n1p] as an example, the precise (translated) text of the messages reads 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!

[19] 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 ensures that reference groups are relatively small – such that drivers have a reasonable chance of earning (avoiding) a positive (negative) message – and avoids indirect treatment interference.

[20] 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.

A printed version of the 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 the targeted messages, so as to account for general feedback effects.[21] The report contains general feedback in the form of a letter score, ranging from A (highest score) to D (lowest score) on the comfort dimensions and fuel economy. Furthermore, it contrasts the overall score of the individual driver with the score of his or her base location. Table 3 summarizes the experimental conditions.

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 about 70% of the drivers receive a negative message in a given feedback round because they score lower than half of their peers on at least one of the three comfort dimensions. In case they perform poorly on multiple dimensions, one is selected randomly for peer-comparison feedback. The remaining 30% performs well on all dimensions and is therefore not notified with 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-month variation in the group of drivers in the top-25% and bottom-50% group. While most drivers move in and out, some drivers are never in the top (bottom) part.[22]

Table 4 summarizes per experimental condition the data in final analysis sample and reports the outcome of balance tests. The p-values show that driver pre-experimental performance in terms of the outcomes fuel economy and ABC events is well-balanced across the experimental groups. A comparison of a rich set of trip-level and bus-type characteristics also reveals no differences across experimental groups, indicating that drivers in

[21] Working with an uninformed control group is not possible due to company policies requiring that every driver should at least receive some 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.

[22] For instance, on acceleration, 19% of the treated drivers is never in the bottom-50% (and 16% always); 42% are never in the top-25% (and 9% always). Outcomes are similar for braking and cornering. Online Appendix D.4 gives a detailed overview per treatment condition of the number of messages send per month and the driving dimensions targeted.

the different treatment groups on average have been exposed to very comparable driving conditions.[23] Drivers are on average 54 years old and work for 20 years at the company. Most drivers are male (89%). The average trip had a length of 31 km and was typically driven in rural areas (84%).

In sum, the detailed data allow for precise identification of good and bad performers in every feedback round. The peer-comparison messages are subsequently intended as a means to assist drivers in offering guidance on where to improve or maintain performance. They are updated every round to inform about progress and to avoid drivers from slacking off. The treatment variation enables us to vary the intensity of the corrective and positive feedback drivers receive.

4.3 In-Person Coaching

In parallel, the company initiated a coaching program. Six experienced drivers (one for each base location) were recruited as coaches based on their track record of driving behavior. All coaches participated in a training on how to approach drivers and how to communicate feedback. Since coaches are bus drivers themselves, there is only limited time available for coaching activities (about one day every two weeks).[24] Furthermore, because of the hop-on hop-off approach to on-the-road coaching, a coach's previous session determines the choice set for the next. This makes random allocation of coaching sessions at the driver-trip level impossible. At the same time, it is next to infeasible for coaches to target specific drivers, also because coaches have no access to the individual feedback reports and hence cannot target drivers with poor scores. We will provide empirical support for the view that the assignment of drivers to coaching is the outcome of a quasi-random process.

In a coaching session, a coach rides along with a bus driver for a portion of the driver's shift. This allows the coach to personalize the feedback and to direct attention to the driver-specific issues at hand. A session is not announced to the driver beforehand. The coach writes down examples of what goes well and wrong and identifies obstacles that may

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

[24]Coaches can decide which day they use for coaching. They vary the day of the week such that every driver has a chance of being coached.

hinder driver performance, such as sharp corners. Due to the presence of passengers, there is no or limited interaction between the driver and the coach during the ride. The coach provides feedback once the trip is completed and passengers have left the bus. The trip is reconstructed using the written-down examples. Both personal and general advice are offered that focus on fuel consumption, punctuality and the ABC dimensions.[25] Drivers are treated as equals and feedback is delivered in a constructive and positive manner.

Coaches maintain a detailed log of their activities, allowing us to pinpoint when and how often drivers are coached. We use these logs to pin down the coaching date. To check whether the assignment of coaching sessions is quasi-random and not based on pre-selected criteria, we compare for each outcome variable (fuel economy and ABC) the mean baseline performance of drivers who have received their first coaching and non-coached colleagues with the same base location. Table 5 verifies balancing on multiple baseline outcome performance measures and covariates. We present both the standard p-values and the ones adjusted for the problem of multiple hypothesis testing using the Bonferroni and Holm correction. Only for morning and evening rush hours we find statistically significant differences. These differences merely reflect that coaches tend to start their work early in the morning. For none of the other variables, we find differences that are even close to significance, especially once we take into account the problem of multiple hypothesis testing. This supports the view that the implementation of the coaching program exhibits a quasi-random order of phase-in.

4.4 Data Collection and Sample Construction

The EOBRs are installed in all three bus types the company operates. The VDL bus is most commonly used, accounting for about 75% of all trips performed. Intouro buses are mainly used for routes with a long travel distance. Two specific features importantly distinguish the IRIS bus from the other bus types. First, both the VDL and Intouro buses have diesel engines, but the IRIS bus runs on natural gas. This implies that for trips completed with an IRIS bus no records on fuel economy are available. Second,

[25] These notes are not included in the logs, so unfortunately we do not know exactly what the coached has conferred with a driver.

whereas the VDL and Intouro buses are used by drivers in all base locations, the IRIS bus is only used by drivers in the largest and most urbanized base location. Hence, the treatment effects on the outcomes fuel economy and ABC events are estimated on the sample of trips completed by either a VDL or Intouro bus.

All 409 tenured drivers are included in our research design. Drivers with a temporary contract, 67 in total, are excluded because their behavior is only observed for short and irregular time spans. The trip-level observations in the final sample are matched with driver, trip, and daily weather characteristics.[26] We use this sample when we analyze the impact of coaching and peer-comparison feedback on a driver's relative ranking. To keep the analysis succinct, we present full estimation results for fuel economy and acceleration in the main text and relegate some findings for braking and cornering events to the online appendix. We will however highlight any important qualitative differences in treatment effects for acceleration and the outcomes braking and cornering when they arise.[27]

5 Results

We first present the results of the written feedback program with the peer-comparison messages (5.1), followed by the effects of the in-person coaching program (5.2). Section 5.3 examines the interference between coaching and the written feedback program.

5.1 Feedback Reports

This section reports the effects of the peer-comparison feedback program. To identify this effect, we estimate the following difference-in-differences (DID) regression specification:

$$\begin{aligned} Y_{its} &= \beta \cdot \text{postannounce}_i + \sum_{j=1}^{4} I\{T_i = j\} \cdot (\tau_j \cdot \text{postfeedback}_{it} + \gamma_j \cdot \text{postexperiment}_{it}) \\ &+ X_{its} \cdot \theta + \mu_i + \kappa_b + \upsilon_t + \zeta_{bt} + \xi_r + \upsilon_{its}. \end{aligned} \tag{1}$$

[26]Online Appendix A details the steps we have taken to construct the final sample.

[27]From the ABC dimensions, we selected acceleration because of the higher average number of events in this dimension (Table 2) and the absence of intermediate changes in threshold settings (Table A2).

The dependent variable Y_{its} is the outcome variable of interest (fuel economy or ABC), indexed by driver (i), time in days (t), and the bus trip (s). In addition, the specification includes a vector X_{its} that contains the control variables listed in Table 1. A rich set of dummy variables controlling for driver (μ_i), bus type (κ_b), day (υ_t), bus type interacted with day (ζ_{bt}), and route (ξ_r) fixed effects completes the specification.[28] Throughout, we use robust standard errors clustered at the driver level to account for within-driver correlation patterns in the error term (Bertrand, Duflo and Mullainathan 2004). Importantly, because coaching takes place in parallel to feedback, a post-coaching dummy variable is included in the controls. In addition, the dummy variable $postfeedback_{it}$ takes on the value one when the first feedback report has been delivered to driver i and is zero otherwise. This definition makes no selection on the actual reading of the report. From a policy perspective, this is useful because it captures the aggregate performance of the treatments when applied to an eligible population (Allcott 2011).[29] The dummy variable $postannounce_{it}$ equals one once a driver is informed about the upcoming EcoManager campaign, and zero otherwise; the dummy $postexperiment_{it}$ equals one once the feedback report with the final notification message has been received, and zero otherwise.

The treatment indicator $T_i = j$ when driver i is assigned treatment j, $j = 1, \ldots, 4$. The τ-coefficients then estimate the treatment-specific effects of receiving tailor-made peer-comparison feedback, while the γ-coefficients measure the impact the withdrawal of peer-comparison messages has on performance (Barankay 2012). The β-coefficient captures the aggregate effect of the launch of the campaign and the switching on of the LED-arrays (which happen at the same date) on driving behavior.

Table 6 presents the results. Our preferred specification is reported in columns (2)-(4) and (6)-(8) and controls for being coached and time-variant driving features, such as weather conditions and the number of passengers. For fuel economy, we find a strong and significant reduction of β=0.41 liters/100km ($0.40\sigma_r$, $p < 0.001$) following the start

[28]By interacting day- and bus type fixed effects, we relax the common trends assumption between bus types to address potential differences over time in the ease (or difficulty) of avoiding ABC events due to different thresholds per bus type. Of course, regressions with day fixed-effects do not include the post announcement and post feedback dummies.

[29]The start of the post-feedback period may differ per driver due to absence in the month on which the first report is based. A no-report indicator captures drivers operating after 15 December 2015 (first feedback round) but who have not yet received their first report.

of EcoManager. This is the joint effect of the launch-event and the switching on of the LED-arrays in the buses. The distribution of feedback reports generates an additional reduction of, on average, 0.13 liters/100km ($0.13\sigma_r$, $p = 0.105$). Column (3) shows that these estimates remain qualitatively unchanged but become more significant once driver fixed effects are added.

How does the experimental variation in the dosage of the number and nature of peer-comparison feedback messages affect worker productivity? Reassuringly, Table 6 shows no differences in fuel economy between the different treatment groups *before* the first feedback report is distributed. The estimates for the post-feedback dummy variable interacted with treatment indicators show no significant effect of the peer-comparison messages in the text boxes in addition to the general effect generated by the feedback reports. The point estimates across treatments for fuel economy are small in size, ranging from -0.11 to 0.05 liters/100km and are individually and jointly insignificant.

While fuel economy is an important outcome variable, the peer-comparison 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. Table 6 reveals that the pattern of effects for acceleration resembles the pattern for fuel economy: a large and significant effect following the announcement ($0.52\sigma_r$), a significant but smaller effect when the feedback reports are received ($0.35\sigma_r$), but again no indication that the text-box variation in the number and nature of peer-comparison messages matters. For braking and cornering, the estimates of the announcement and reception of feedback are $1.23\sigma_r$, $0.00\sigma_r$, and $0.14\sigma_r$, $0.10\sigma_r$, respectively (Table A5).[30] Also for these dimensions, we do not observe any peer-comparison effect.

In sum, with the exception of braking, the launch of the feedback program and the distribution has a significantly positive impact on fuel economy and all ABC dimensions. Table A14 shows that results remain significant at the $p = 0.05$-level when we apply a

[30]The larger effect on braking is partly due to a change in the threshold setting for braking-events happening around the same time, see Appendix A.1.

Holm correction for multiple hypothesis testing.[31]

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 sub-groups 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.[32] There is no indication that drivers respond differently to the first peer-comparison messages that they receive than to the ones received in later months, for example because they lose attention. What about the possibility that certain sub-groups of drivers are more responsive to the peer-comparison feedback program than others? Given our design feature that only the sub-set of drivers who actually belong to the top-25% or bottom-50% receive a message, it is indeed possible that we overlook some treatment effects among the subgroups that are treated. 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. For example, every month only about 70% of all drivers in treatments T2[1n0p] and T4[3n0p] actually receive messages in their textbox.[33]

In an explorative analysis, we group drivers on basis of their performance in month m (being in the top-25% or bottom-50%) and (for each group and ABC outcome dimension separately) regress a driver's ranking in month $m+1$ on a dummy variable on receiving relative performance feedback on month m performance. The coefficient estimates (reported in Table A10) for cornering all are insignificant. For acceleration and braking, the coefficients for drivers in the bottom-50% are negative and in some cases significant, suggesting that the feedback messages help them to improve their ranking; for drivers in

[31] At first sight, the fuel economy estimates for 'post-experiment' seem to suggest that the withdrawal of peer-comparison messages in the text box completely reverses the improvement in fuel economy achieved by the introduction of the EcoManager program: $|0.554| \sim |0.411 + 0.130|$. However, caution is needed in drawing this conclusion because the post-experimental period is relatively short and the specifications lack day fixed effects to absorb the unobserved day-to-day fluctuations in driving conditions. Also, the post-experiment coefficients for the ABC outcomes do not reflect a rebound effect.

[32] Appendix G.1 contains the coefficient plots of these estimates.

[33] 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 negative feedback (54%), only positive feedback (25%) or a combination of positive and negative feedback (21%). See Table A9 for detailed information on the composition of messages by treatment.

the top-25%, the coefficients are consistently and significantly positive, indicating that their average ranking deteriorates when having received positive feedback.

5.2 In-Person Coaching

To identify the effect of a single on-the-road coaching session on productivity outcomes in the weeks following coaching, we estimate the following DID regression specification:

$$\begin{aligned} Y_{its} &= \delta_0 I\{t = t_i^c\} + \sum_{\tau=1}^{10} \delta_\tau I\left\{t - t_i^c \in (7(\tau-1), 7\tau]\right\} + \delta_{10}^{+} I\left\{t - t_i^c > 70\right\} \qquad (2) \\ &+ \sum_{\tau=1}^{10} \gamma_{-\tau} I\{t - t_i^c \in [-7\tau, -7(\tau-1))\} + X_{its} \cdot \theta + \mu_i + \kappa_b + \upsilon_t + \zeta_{bt} + \xi_r + \epsilon_{its}. \end{aligned}$$

As before, the dependent variable Y_{its} denotes the outcome of interest, the same set of control variables as in equation (2) is included and standard errors are clustered at the driver level. Day t_i^c denotes the specific day at which driver i is coached; recall that because of the phase-in design of the coaching program, drivers are coached at different days. The regressors include indicator functions $I(\cdot)$ to estimate the impact of coaching at: *a*) the day of coaching (coefficient δ_0); *b*) the first ten weeks following coaching (post-coaching coefficients $\delta_1, \ldots, \delta_{10}$), and *c*) the ten weeks preceding coaching (pre-coaching coefficients $\gamma_{-1}, \ldots, \gamma_{-10}$). The coefficient δ_{10}^{+} absorbs any impact of coaching more than 10 weeks after the day of coaching.[34]

Figure 2 shows the temporal effects of coaching by plotting the pre- and post-coaching coefficients effects for the fuel economy and ABC outcomes. For fuel economy and acceleration, we observe a strong and immediate effect of coaching: on the day of coaching the fuel need reduces by 0.6 liters/100km ($0.58\sigma_r$) and the number of acceleration events by 1.1 events/10km ($0.50\sigma_r$). These effect sizes are respectively about 1.5 and 1.0 times the impact of the start of EcoManager. These effects persist for about seven to nine weeks. This suggests that, as time progresses, coaching effects decay and drivers seem to fall back into old driving habits.[35] We also observe an effect of coaching for braking and cornering,

[34] Observations more than 10 weeks before coaching are the omitted period, the estimated δ-coefficients in Figure 2 show the average effect relative to this baseline period.

[35] This fits into the body of evidence showing that in many domains, it is hard to induce persistent

but these effects are much less pronounced and not (braking) or less (cornering) significant because of the lower baseline number of events (see Table 2). For none of the outcomes we observe differences in driving behavior in the 10 weeks prior to coaching, which lends support to our earlier conclusion that the selection for a coaching session is quasi-random and not based on prior performance.

Table 7 reports the main coefficients of regression specifications that take the entire period preceding coaching as the baseline period. Next to the standard p-values, we also report p-values that apply a Bonferroni and a Holm correction for multiple hypothesis testing (MHT). These are conservative methods to adjust for the fact that we consider the impact of coaching on four different outcome variables and three different time periods.[36]

The regressions reveal that the largest improvements are observed on the day of coaching and with all adjusted p-values < 0.02: fuel economy improves by 0.61 liters/100km ($0.55\sigma_r$), acceleration, braking and cornering by $0.48\sigma_r$, $0.11\sigma_r$ and $0.10\sigma_r$, respectively. For all outcomes except braking, we identify a short-run persistence effect in the first week following coaching. Only for acceleration, an effect is identified for the entire post-coaching period.

5.2.1 Robustness Check: Heterogeneity in Coach Quality

Coaching is provided by a small number of six coaches. One potential worry then is that the observed average treatment effects are not caused by an inherent feature of in-person coaching that is independent of who coaches, but is instead due to one or two coaches with idiosyncratic coaching qualities that are hard to copy. In that case, the data would not allow us to draw the general conclusion that in-person coaching improves worker productivity. We cannot directly compare differences in the way our six coaches provided feedback to drivers because we lack this information. We can however estimate the treat-

changes in habits (Brandon, Ferraro, List, Metcalfe, Price and Rundhammer 2017).

[36]The Bonferroni multiplicity-adjusted p-values are obtained by multiplying the unadjusted p-values by the number of hypotheses (12); the Holm multiplicity-adjusted p-values are obtained by ranking the unadjusted p-values from largest to smallest and to multiply each unadjusted p-value with its rank. Due to our stratified design, we cannot apply the less conservative MHT correction method developed by List, Shaikh and Xu (2019) that assumes simple random matching. When the joint dependence between the individual test statistics is positive (which is likely in our case given the positive correlations in Table 2) the latter method has a greater ability to detect false null hypotheses than the Bonferroni and Holm method.

ment effect of each individual coach by considering the sub-sample of drivers instructed by that coach. When there is substantial heterogeneity in the quality of instructions given by the coaches, this should result in between-coach differences in treatment effects.

Figure 3 shows these coach-level treatment effects of in-person coaching for the outcome fuel economy.[37] Despite the fact that the estimates are less precise due to the smaller sub-samples, the pattern is remarkably consistent across coaches: on the day of coaching, for all coaches the point estimate of fuel savings is in the range [-0.7, -0.4] liters/100km.[38] The diminishing and eventually vanishing of this effect in the seven to nine weeks following coaching is common to all coaches. Based on this evidence, we conclude that the observed effect can be attributed to features inherent to in-person coaching.

5.2.2 Treatment Heterogeneity

Next we address whether there is heterogeneity in driver responses to coaching. The literature on peer effects in educational outcomes suggests that the effect of coaching may be heterogeneous, depending on a driver's own past performance. In this section we address the open question whether this result carries over to non-educational contexts. We take the following non-parametric approach. For a driver coached in month m, we compare the driver's relative performance in productivity outcome y the month before $(m-1)$ and the month $(m+1)$ following coaching. We thus ignore a driver's relative performance in the month in which (s)he has been coached. We do this for all four productivity measures. Of course, because of reversion to the mean, there is a tendency for drivers who by chance attain a particularly high (low) ranking in month $m-1$ to have a lower (higher) ranking in month $m+1$. To account for this statistical phenomenon, we use the change in ranking non-coached drivers experience from month $m-1$ to $m+1$ as a benchmark against we evaluate the change in ranking of drivers coached in month m.

Figure 4 plots for both groups the change in ranking. For non-coached drivers, the shaded area represents the local polynomial estimates of the relation between the ranking

[37]Figures A6-A8 in the appendix show the coach-level treatment effects for the ABC dimensions.

[38]Estimates for coach # 3 are ignored. These estimates are very imprecise because this coach operates in the urban area with IRIS buses for which fuel economy is not recorded.

in months $m-1$ and $m+1$, along with a 95% confidence interval.[39] We fit separate polynomials for non-coached drivers part of the top-25%/bottom-50%/remaining group in month $m-1$. Due to the reversion to the mean effect, the slope of each of these polynomials is less than one. The plots show clear evidence of heterogeneity in the effects of coaching: only drivers at the bottom half of the performance distribution benefit from coaching.[40] This result holds independent of which productivity outcome is considered (fuel economy or either of the comfort dimensions). The direction of our result is in contrast to the empirical literature on peer effects in education, which predominantly finds that high-achieving students benefit most from the presence of high-achieving peers. Possible explanations for this difference are that high-performing workers in our setting have little room left for further improvement or are less open to a colleague's feedback.

5.3 Treatment Interaction

We conclude with an exploratory analysis on the possible complementarity between in-person coaching and peer-comparison feedback. For this, we utilize the fact that a sub-set of drivers received coaching before receiving written feedback while others received one or more written feedback reports before being coached.

We first consider whether having received the general feedback reports makes drivers more or less responsive to coaching. For fuel economy and acceleration, Figure 5 compares the response to coaching by drivers who did not yet receive feedback on paper with those who did. Although the confidence intervals have become wider because of the smaller samples, a comparison of panels (*a*) with (*b*) and (*c*) with (*d*) shows for both groups a similar pattern in the effect of coaching, both on the day of coaching as well as in the subsequent weeks. Hence, the effect of in-person coaching does not depend on having received prior feedback on one's performance in written form. We also checked whether the impact of coaching is affected by the treatment variation in the number and nature of peer-comparison messages in the tex box. In line with the non-significant effects of the variation in text-box messages discussed in Section 5.1, we find no effect (see Appendix H).

[39] In calculating the weighted local estimate, we use the standard Epanechnikov kernel function.
[40] See Table A13 for regression estimates.

What about the opposite case: do drivers who did already receive in-person coaching respond differently to the peer-comparison feedback messages than those who did not yet receive coaching? To answer this question, we compare the response to the feedback reports by drivers who were coached before the arrival date of the first report (December 15, 2015) with the response by drivers who did not receive coaching at all. For both subsamples, we run the same regression specification as in the previous section. Table 8 shows the results. Coached drivers seem more responsive to the general feedback report ('post-feedback'). Of most interest is the difference in response to the peer-comparison text-box messages. The treatment variation in the number of peer-comparison messages does not generate any observable change in productivity among the group of coached drivers, similar to what we found earlier for the entire sample. However, in the group of drivers that has not yet been exposed to coaching, varying the nature and intensity of feedback does seem to have an effect. Drivers in the treatment group exposed to the highest number of negative feedback messages [3n0p] improve significantly in acceleration (p=0.003), braking (p=0.002) and fuel economy (p=0.041) compared to non-coached drivers that do not receive any peer-comparison messages. For fuel economy and braking, we also find positive effects in the group that is exposed to up to one negative message [1n0p] but at lower levels of significance (p=0.016 and p=0.077, respectively). For none of the outcome variables we find a treatment effect for treatment [1n1p] that mixes negative and positive feedback.

In sum, in-person coaching and peer-comparison feedback seem to interfere in an asymmetric manner: coaching is effective independent of prior exposure to peer-comparison messages but prior coaching renders peer-comparison messages non-effective. One possible explanation is that in-person coaching trumps peer-comparison feedback: once drivers have met a coach who gave them detailed feedback on what they do right and wrong on a trip, they become insensitive to subsequent messages about their relative performance. Our evidence shows no need to limit negative feedback or to mix it with positive feedback.

6 Discussion and Conclusions

Given our precise empirical estimates on the impact of the written feedback and the in-person coaching program on worker productivity, we next discuss the possible channels through which these programs change drivers' behavior. Cassar and Meier (2018) present a theoretical framework in which they distinguish four factors that affect work meaning: the need for autonomy, competence and relatedness and the mission of an organization.[41] Different features of the feedback programs may impact these four dimensions of work meaning. The announcement of EcoManager and the provision of feedback may help to align drivers' beliefs with the (social) mission of the firm. Corrective peer-comparison feedback can help to develop competence, but may also make a driver feel less competent. To avoid the latter, it may work to combine corrective feedback with positive feedback. Intensifying feedback may on the other hand also induce feelings of being monitored and a loss in autonomy. Finally, being coached by an experienced colleague may strengthen the social relation with colleagues, thus benefiting relatedness.

The announcement of EcoManager has a strong and positive impact on all four productivity measures. One possible channel is that the campaign makes the social mission of the firm salient, thereby increasing the workers sense of job meaning. Similar effects have been recorded in fundraising contexts by Grant (2008). From a principal-agent perspective, another possibility is that the announcement triggers reputational concerns – despite the firm's assurance that the feedback information will not be used in formal evaluations (List 2003). In line with evidence that checklists improve worker productivity by serving as a "memory aid" (Jackson and Schneider 2015), it is also possible that the switching on of the LED-arrays in the bus, happening around the feedback announcement date, serves as a permanent memory aid to drive carefully.

Additional regressions show that the announcement effect on fuel economy and acceleration for drivers less than fifty years old is about twice the size of that of other drivers and highly significant.[42] Figure 6 illustrates this for fuel economy. The lines show per

[41]The first three are psychological needs that have been identified by self-determination theory as essential for human motivation, see Ryan and Deci (2000) and references therein.

[42]The outcome variables are regressed on post-announcement, post-feedback and post-experiment dummies, each interacted with four about equally sized age-categories (aged < 50, $50 - 54$, $55 - 59$ or ≥ 60).

age, years-of-service or gender category, respectively, the week-dummy coefficients over time (the week before the announcement, week 40, is taken as the baseline). Clearly, younger drivers show a sharper response to the EcoManager launch and the gap with the older drivers never closes (panel (a)). Does this reflect cohort differences in learning or reputational concerns? In case of the latter, we might expect a similar gap to occur if we categorize drivers by the number of years that they are already at the company. We do not observe this (panel (b)), suggesting that the increased saliency of the company's objectives especially resonates with younger drivers.

We find a positive effect of receiving written feedback. The feedback may indeed help drivers to become more competent drivers. Of course, drivers may do better on the dimensions measured because they know that these are monitored by the company. There is the possibility that their performance on unmeasured dimensions of job performance such as friendliness will deteriorate. We have no information on this but also did not hear from the company that the number of complaints increased. No treatment effects for the peer-comparisons in the text-messages are identified, except that not-yet-coached drivers show a larger improvement if they receive the full amount of corrective feedback. When workers receive negative feedback on certain dimensions, it does not increase their performance on these dimensions if the corrective feedback is accompanied by positive feedback on other dimensions.

Different explanations are possible for the strong and immediate impact of coaching. Coaching may improve the driver's competences, improve the worker's alignment with the company's mission, and/or deepen the relatedness to colleagues. We cannot distinguish between these three. However, the decay path tells us that neither the improved alignment with firm objectives, nor the improvement to human capital, nor the closer connection to colleagues is permanent. A social pressure explanation does not fit the pattern because the decay is not immediate. It may be that coaching serves as a memory aid for drivers with limited attention. Hanna, Mullainathan and Schwartzstein (2014) find a similar result for farmers. Farmers change behavior "when presented with summaries that highlight previously unattended-to relationships in the data". Their paper cannot tell whether this

Full regression estimates reported in Tables A20-A22.

effect remains as their final follow-up is two months after farmers receive this information. Our evidence clearly points out that, at least for bus drivers, the effect is short-lived.

Comparing the written feedback and in-person feedback, it seems that in-person feedback is more powerful but that the repetitive character of the monthly written report is necessary to let the impact last. This is in line with Dusch, Evans, Eze-Ajoku and Macis (2017) and also resonates the recommendations of Oreopoulos and Petronijevic (2018) for feedback provision in educational settings.

To conclude, increasing conservation efforts among individuals is generally seen as low-hanging fruit in the battle to reduce energy consumption. Most existing studies on energy conservation however focus on households. Their results may not apply in corporate settings where workers are in a principal-agent relationship and have no financial stake in energy conservation and where institutional constraints often hinder firms in the use of pay-for-performance plans. Our findings contribute to a growing body of literature on non-financial interventions aimed at energy conservation in the workplace. Our analyses show how carefully designed non-pecuniary strategies can improve conservation efforts among workers. One robust outcome of the quasi-experimental variation in being coached is that in-person coaching immediately improves a driver's performance. Especially drivers with the lowest prior performance receive a boost. Although this effect is transient, the decay is not immediate. In the weeks following coaching, 17.52 liter of fuel is saved per coached driver, which amounts to €19.27, or €60 per day of coaching. This is less than the cost of freeing up an experienced driver to coach, but the benefits may outweigh the cost if the company or passengers sufficiently value the improved comfort or environmental gains. Written performance feedback comes at a lower marginal cost. The reduction in fuel consumption following the announcement and distribution of general feedback reports is 5/6th the effect of coaching on the day of coaching. The additional peer-comparison messages do not have any impact and we find no evidence that managers need to restrain themselves from giving negative feedback.

Our research points to several directions for future research. First, the increased adoption of digital monitoring technologies at the work floor creates a myriad of new opportunities for tailoring feedback to motivate workers. This paper has examined the impact of different feedback channels and variation in feedback intensity, but we reckon that designing and evaluating other data-driven incentives could yield fruitful research. Second, we document important interaction effects between written feedback and in-person coaching and believe that more research should be done to investigate interactions between non-financial incentives. Finally, and more in general, the question how conservation efforts can be stimulated when someone else pays the bill is in need of more answers.

References

Allcott, Hunt, "Social Norms and Energy Conservation," *Journal of Public Economics*, 2011, *95* (9-10), 1082–1095.

___ **and Judd B. Kessler**, "The Welfare Effects of Nudges: A Case Study of Energy Use Social Comparisons," *American Economic Journal: Applied Economics*, 2019, *11* (1), 236–276.

___ **and Sendhil Mullainathan**, "Behavior and Energy Policy," *Science*, 2010, *327* (5970), 1204–1205.

___ **and Todd Rogers**, "The Short-Run and Long-Run Effects of Behavioral Interventions: Experimental Evidence from Energy Conservation," *American Economic Review*, 2014, *104* (10), 3003–3037.

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.

Ayres, Ian, Sophie Rasemand, and Alice Shih, "Evidence from Two Large Field Experiments that Peer Comparison Feedback Can Reduce Residential Energy Usage," *Journal of Law, Economics, and Organization*, October 2013, *29* (5), 992–1022.

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.

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

___ **and Thomas N. Hubbard**, "Make Versus Buy in Trucking: Asset Ownership, Job Design, and Information," *American Economic Review*, 2003, *93* (3), 551–572.

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.

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

Barkenbus, Jack N., "Eco-driving: An overlooked climate change initiative," *Energy Policy*, 2010, *30*, 762–769.

Benabou, Roland and Jean Tirole, "Incentives and Prosocial Behavior," *American Economic Review*, 2006, *96* (5), 1652–1678.

Bertrand, Marianne, Esther Duflo, and Sendhil Mullainathan, "How Much Should We Trust Differences-in-Differences Estimates," *The Quarterly Journal of Economics*, 2004, *119* (1), 249–275.

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.

Booij, Adam S., Edwin Leuven, and Hessel Oosterbeek, "Ability Peer Effects in University: Evidence from a Randomized Experiment," *Review of Economic Studies*, 2017, *84*, 547–578.

Brandon, Alec, Paul J. Ferraro, John A. List, Robert D. Metcalfe, Michael K. Price, and Florian Rundhammer, "Do The Effects of Social Nudges Persist? Theory and Evidence from 38 Natural Field Experiments," NBER working paper no. 23277 March 2017.

Brynjolfsson, Erik. and Kristina McElheran, "The Rapid Adoption of Data-Driven Decision-Making," *American Economic Review*, 2016, *106* (5), 133–139.

Burke, Mary A. and Tim R. Sass, "Classroom Peer Effects and Student Achievement," *Journal of Labor Economics*, 2013, *31* (1), 51–82.

Cassar, Lea and Stephan Meier, "Nonmonetary Incentives and the Implications of Work as a Source of Meaning," *Journal of Economic Perspectives*, Summer 2018, *32* (3), 215–238.

Chetty, Raj, John N. Friedman, and Jonah E. Rockoff, "Measuring the Impacts of Teachers I: Evaluating Bias in Teacher Value-Added Estimates," *American Economic Review*, September 2014, *104* (9), 2593–2632.

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).

Dunsch, Felipe A., David K. Evans, Ezinne Eze-Ajoku, and Mario Macis, "Management, supervision, and healthcare: A field experiment," NBER working paper no. 23749 August 2017.

Edmunds, Angela and Anne Morris, "The problem of information overload in business organisations: a review of the literature," *International Journal of Information Management*, 2000, *20*, 17–28.

Eriksson, Tor, Anders Poulsen, and Marie Claire Villeval, "Feedback and Incentives: Experimental Evidence," *Labour Economics*, 2009, *16* (6), 679–688.

Ferraro, Paul J. and Michael K. Price, "Using Nonpecuniary Strategies to Influence Behavior: Evidence from a Large-Scale Field Experiment," *The Review of Economics and Statistics*, 2013, *95* (1), 64–73.

Freeman, Richard B., "Union Wage Practices and Wage Dispersion within Establishments," *Industrial and Labor Relations Review*, 1981, *36* (1), 3–21.

Fruehwirth, Jane Cooley, "Identifying peer achievement spillovers: Implications for desegregation and the achievement gap," *Quantitative Economics*, 2013, *4*, 85–124.

Gerarden, Todd D., Richard G. Newell, and Robert N. Stavins, "Assessing the Energy-Efficiency Gap," *Journal of Economic Literature*, December 2017, *55* (4), 1486–1525.

Golman, Russell, David Hagmann, and George Loewenstein, "Information Avoidance," *Journal of Economic Literature*, 2017, *55* (1), 96–135.

Gosnell, Greer K., John. A List, and Robert Metcalfe, "The Impact of Management Practices on Employee Productivity: A Field Experiment with Airline Captains," *Journal of Political Economy*, April 2020, *128* (4), 1195–1233.

Grant, Adam M., "The Significance of Task Significance: Job Performance Effects, Relational Mechanisms, and Boundary Conditions," *Journal of Applied Psychology*, 2008, *93* (1), 108–124.

Hahn, Robert., Robert D. Metcalfe, David Novgorodsky, and Michael K. Price, "The Behavioralist as Policy Designer: The Need to Test Multiple Treatments to Meet Multiple Targets," *NBER Working Paper Series.*, 2016, (No. 22886).

Hanna, Rema, Sendhil Mullainathan, and Joshua Schwartzstein., "Learning Through Noticing: Theory and Experimental Evidence in Farming," *Quarterly Journal of Economics*, 2014, *129* (3), 1311–1353.

Hitt, Lorin M. and Erik Brynjolfsson, "Information Technology and Internal Firm Organization: An Exploratory Analysis," *Journal of Management Information Systems*, Fall 1997, *14* (2), 81–101.

Holladay, Scott J., Jacob LaRiviere, David M. Novgorodsky, and Michael K. Price, "Asymmetric Effects of Non-pecuniary Signals on Search and Purchase Behavior for Energy-Efficient Durable Goods," *NBER Working Paper Series*, 2016, (No. 22939).

Hölmstrom, Bengt, "Moral Hazard and Observability," *The Bell Journal of Economics*, 1979, *10* (1), 74–91.

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

Hoxby, Caroline M. and Gretchen Weingarth, "Taking race out of the equation: School reassignment and the structure of peer effects," `https://www.pausd.org/sites/default/files/pdf-faqs/attachments/TakingRaceOutOfTheEquation.pdf` 2005.

Hubbard, Thomas N., "Information, Decisions, and Productivity: On-Board Computers and Capacity Utilization in Trucking," *American Economic Review*, 2003, *93* (4), 1328–1353.

ICCT, "From Laboratory to Road: A Comparison of Official and 'Real-World' Fuel Consumption CO2 Values for Cars in Europe and United States," Technical Report, International Council on Clean Transportation 2013.

IEA, "Improving the Fuel Economy of Road Vehicles: A Policy Package," Technical Report, International Energy Agency 2012.

Jackson, C. Kirabo and Henry S. Schneider, "Checklists and Worker Behavior: A Field Experiment," *American Economic Journal: Applied Economics*, 2015, *7* (4), 136–168.

Kelley, Erin M., Gregory Lane, and David Schönholzer, "The Impact of Monitoring Technologies on Contracts and Employee Behavior: Experimental Evidence from Kenya's Transit Industry," mimeo, University of Berkeley 2018.

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

Lam, Chak Fu, D. Scott DeRue, Elizabeth P. Karam, and John R. Hollenbeck, "The impact of feedback frequency on learning and task performance: challenging the "more is better" assumption," *Organizational Behavior and Human Decision Processes*, 2011, *116* (2), 217–228.

Larrick, Richard P. and Jack B. Soll, "The MPG Illusion," *Science*, 2008, *320* (5883), 1593–1594.

Lavy, Victor, M. Daniele Paserman, and Analia Schlosser, "Inside the black box of ability peer effects: Evidence from variation in the proportion of low achievers in the classroom," *The Economic Journal*, March 2011, *122*, 208–237.

——**, Olmo Silva, and Felix Weinhardt**, "The Good, the Bad, and the Average: Evidence on Ability Peer Effects in Schools," *Journal of Labor Economics*, 2012, *30* (2), 367–414.

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

List, Johan A., Azeem M. Shaikh, and Yang Xu, "Multiple hypothesis testing in experimental economics," *Experimental Economics*, 2019, *22*, 773–793.

List, John A., "The Behavioralist Meets the Market: Measuring Social Preferences and Reputation Effects in Actual Transactions," *Journal of Political Economy*, 2003, *2006* (1), 1–37.

MIVW, "Public Transport in the Netherlands," Technical Report, Ministry of Transport, Public Works and Water Management, `https://www.emta.com/IMG/pdf/brochure.pdf` June 2010.

Moldovanu, Benny, Aner Sela, and Xianwen Shi, "Contests for Status," *Journal of Political Economy*, 2007, *115* (2), 338–363.

Nilekani, Janhavi, "Driving Down Demand for Diesel: Does a Bus Driver Training and Incentive Program Increase Fuel Efficiency?," working paper January 6 2018.

Oreopoulos, Philip and Uros Petronijevic, "Student Coaching: How Far Can Technology Go?," *Journal of Human Resources*, Spring 2018, *53* (2), 299–329.

Pierce, Lamar, Daniel C. Snow, and Andrew McAfee, "Cleaning House: The Impact of Information Technology Monitoring on Employee Theft and Productivity," *Management Science*, October 2015, *61* (10), 2299–2319.

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.

Sacerdote, Bruce, "Peer Effects in Education: How Might They Work, How Big Are They and How Much Do We Know Thus Far?," in Ludger Woessmann Eric A. Hanushek, Stephen Machin, ed., *Handbook of the Economics of Education*, Vol. 3, Elsevier, 2011, chapter 4, pp. 249–277.

Sandvik, Jason, Richard Saouma, Nathan Seegert, and Christopher Stanton, "Workplace Knowledge Flows," CEPR Discussion Paper No. DP14299. Available at SSRN: `https://ssrn.com/abstract=3518641` January 2020.

Simon, Herbert A., "Applying Information Technology to Organization Design," *Public Administration Review*, May-June 1973, *33* (3), 268–278.

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: Determinants of Fuel Economy

Dependent variable:	Fuel economy				
	(1)	(2)	(3)	(4)	(5)
VDL 10m	-1.013***	-1.014***	-0.948***	-0.943***	-0.945***
	(0.029)	(0.029)	(0.028)	(0.027)	(0.016)
VDL 14m	8.910***	8.958***	8.904***	8.935***	8.922***
	(0.165)	(0.158)	(0.145)	(0.146)	(0.136)
Intouro	4.785***	4.808***	4.445***	4.487***	4.483***
	(0.239)	(0.239)	(0.234)	(0.216)	(0.161)
Rush hour 7-10am			-0.349***	-0.350***	-0.248***
			(0.031)	(0.030)	(0.015)
Rush hour 4-7pm			0.334***	0.334***	0.246***
			(0.039)	(0.038)	(0.018)
Non-scheduled trip			0.095	0.077	0.096**
			(0.093)	(0.091)	(0.044)
No. of stops per km.			0.721***	0.750***	0.715***
			(0.087)	(0.086)	(0.083)
Urban trip			0.000	0.278	0.673
			(.)	(1.169)	(0.838)
Trip length (in km.)			-0.023***	-0.025***	-0.024***
			(0.001)	(0.001)	(0.001)
Ln(No. of passengers)			1.263***	1.288***	1.275***
			(0.016)	(0.016)	(0.014)
Punctuality			0.013***	0.010**	0.045***
			(0.005)	(0.005)	(0.003)
Constant	23.716***	23.450***	20.531***	21.274***	21.156***
	(0.053)	(0.057)	(0.134)	(0.182)	(0.166)
R^2	.279	.291	.409	.43	.514
Number of trip-level observations	533171	533171	533171	533171	533171
Weather dummies	No	Yes	Yes	No	No
Driver fixed effects	No	No	No	No	Yes
Day fixed effects	No	No	No	Yes	Yes
Route fixed effects	No	No	Yes	Yes	Yes

Notes: Dependent variable: Fuel economy in liters/100km. Default categories: Scheduled trips in rural areas outside rush hours completed with a VDL 12m bus. Trip punctuality is the difference in minutes between actual and scheduled driving time. Weather data are collected from a weather station located in the regional capital and are maintained by the Royal Netherlands Meteorological Institute (KNMI). Standard errors are clustered by driver.
***(**,*) : statistically different from zero at the 1%-level (5%-level, 10%-level). Standard errors in parentheses.

Table 2: Driver Variation in Fuel Economy and ABC Comfort Outcomes

					Residual correlation		
Driving dimension	μ	σ_r	c_v	$\Delta(p_{90} - p_{10})$	Acc.	Braking	Corn.
Fuel economy	24.91	1.03	0.04	2.46	0.399	0.083	0.166
Acceleration	10.87	2.22	0.20	4.87		0.268	0.271
Braking	1.68	1.02	0.61	0.93			0.154
Cornering	2.27	1.93	0.85	1.71			

Notes: Fuel economy: liters/100km. ABC dimensions: no. of events/10km. σ_r: residual standard deviation; c_v: coefficient of variation. $\Delta(p_{90} - p_{10})$: difference in performance 90th vs. 10th percentile driver.

Table 3: Experimental Conditions

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

Table 4: 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 effect = 0	
	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.		
	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 7:00-10:00 and 16:00-19:00, respectively. Holiday rides take place during, for example, Christmas, New Year's Eve and school holidays. 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 5: Quasi-Random Coaching: Balancing Tests on Baseline Performance and Non-Performance Descriptives

	C	NC	Δ(C-NC)	stand. p-value	Bonf. corr.	Holm corr.
Baseline performance						
Fuel economy	25.038	25.247	-0.209	0.0865	1.0000	0.2595
Acceleration	13.132	13.349	-0.217	0.5200	1.0000	1.0000
Braking	3.715	3.789	-0.075	0.7744	1.0000	1.0000
Cornering	1.093	1.189	-0.096	0.2065	1.0000	1.0000
Share of experimental conditions						
T1 (0n0p)	0.260	0.247	0.013	0.8118	1.0000	1.0000
T2 (1n0p)	0.244	0.261	-0.016	0.7602	1.0000	1.0000
T3 (1n1p)	0.235	0.228	0.007	0.8968	1.0000	1.0000
T4 (3n0p)	0.260	0.264	-0.003	0.9532	1.0000	1.0000
Demographics						
Year of birth	1962.4	1961.8	0.558	0.3372	1.0000	1.0000
Year of employment	1996.2	1996.3	-0.150	0.8706	1.0000	1.0000
Share of FTE≥0.9	0.801	0.776	0.025	0.6205	1.0000	1.0000
Share of female drivers	0.094	0.073	0.020	0.554	1.0000	1.0000
Trip-specific variables						
Punctuality	-2.942	-3.018	0.076	0.2251	1.0000	1.0000
Distance traveled (in km.)	30.564	32.025	-1.461	0.0996*	1.0000	0.3984
Number of passengers	15.035	15.283	-0.248	0.5983	1.0000	1.0000
Number of bus stops	37.934	37.864	0.070	0.9268	1.0000	1.0000
Share of rides:						
- Morning rush hours	0.298	0.159	0.139	0.0076***	0.1976	0.0076***
- Evening rush hours	0.125	0.253	-0.127	0.0084***	0.2184	0.0168**
- Weekend	0.031	0.031	0.000	1.0000	1.0000	1.0000
- Fill in	0.004	0.025	-0.021	0.1539	1.0000	0.7695
- Holidays	0.119	0.125	-0.006	0.8839	1.0000	1.0000
- Urban area	0.149	0.146	0.003	0.9366	1.0000	1.0000
- School	0.004	0.004	0.000	0.9954	1.0000	1.0000
Share of rides on bus types						
VDL	0.773	0.755	0.017	0.7427	1.0000	1.0000
Intouro	0.078	0.099	-0.021	0.5555	1.0000	1.0000
IRIS	0.149	0.146	0.003	0.9366	1.0000	1.0000

Notes: Unit of observation is the driver. For every coaching date, the mean baseline performance and non-performance related variables of drivers who receive their first coaching (C) is compared to that of non-yet-coached colleagues (NC). Reported are the mean values over all coaching dates. Standard p-values as well as p-values that use a Bonferroni and Holm correction for multiple hypothesis testing are reported. Fuel economy in liters/100km; Performance on the ABC dimensions as the number of events/10km. Fewer events mean better driving behavior. Punctuality is the difference in minutes between actual and planned driving time. Number of passengers is based on check-ins with public transport cards. Morning and evening rush hours are from 7:00-10:00 and 16:00-19:00, respectively. Holiday rides take place during, for example, Christmas, New Year's Eve and school holidays. School rides are along routes with schools and universities as final destinations. Fill-ins are non-scheduled trips whereby a driver replaces a colleague from another base location.

$^{***}(^{**},^{*})$: the corresponding p-values are less than 1% (5% or 10%).

Table 6: Targeted Peer-Comparison Feedback Effects on Driving Performance

Dependent variable:	Fuel Economy				Acceleration			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Post-announcement	-0.305***	-0.411***	-0.367***		-2.250***	-1.154***	-1.229***	
	(0.062)	(0.053)	(0.038)		(0.129)	(0.116)	(0.094)	
T2 (1n/0p)	-0.258*	-0.201			-0.328	-0.348		
	(0.154)	(0.125)			(0.277)	(0.249)		
T3 (1n/1p)	-0.040	-0.079			0.160	0.025		
	(0.172)	(0.152)			(0.362)	(0.305)		
T4 (3n/0p)	-0.081	-0.121			0.243	0.167		
	(0.164)	(0.134)			(0.284)	(0.256)		
Post-feedback	-0.402***	-0.130	-0.173***		-0.656***	-0.824***	-0.693***	
	(0.077)	(0.080)	(0.062)		(0.159)	(0.168)	(0.142)	
Post-feedback × T2 (1n/0p)	-0.105	-0.090	-0.026	-0.026	0.045	0.100	-0.097	-0.110
	(0.116)	(0.092)	(0.083)	(0.082)	(0.238)	(0.216)	(0.218)	(0.215)
Post-feedback × T3 (1n/1p)	-0.013	0.047	0.036	0.023	-0.061	0.083	0.027	0.039
	(0.096)	(0.079)	(0.080)	(0.078)	(0.227)	(0.200)	(0.199)	(0.196)
Post-feedback × T4 (3n/0p)	-0.082	0.017	-0.021	-0.032	-0.237	-0.117	-0.252	-0.236
	(0.103)	(0.075)	(0.074)	(0.072)	(0.231)	(0.204)	(0.208)	(0.207)
Post-experiment	0.574***	0.554***	0.553***		0.367**	0.142	0.187*	
	(0.089)	(0.062)	(0.055)		(0.142)	(0.119)	(0.104)	
Post-experiment × T2 (1n/0p)	-0.010	0.018	0.033	0.036	0.114	0.016	0.032	0.029
	(0.130)	(0.093)	(0.079)	(0.077)	(0.220)	(0.187)	(0.149)	(0.149)
Post-experiment × T3 (1n/1p)	0.048	0.070	0.027	0.040	0.081	0.151	0.065	0.055
	(0.114)	(0.081)	(0.070)	(0.069)	(0.188)	(0.158)	(0.137)	(0.138)
Post-experiment × T4 (3n/0p)	-0.056	0.050	0.033	0.023	0.017	0.157	0.080	0.060
	(0.115)	(0.084)	(0.079)	(0.076)	(0.198)	(0.162)	(0.140)	(0.140)
Constant	24.639***	22.056***	22.546***	23.984***	10.790***	8.594***	11.868***	11.863***
	(0.110)	(0.104)	(0.080)	(0.181)	(0.217)	(0.194)	(0.191)	(0.259)
R^2	.0148	.417	.499	.515	.0527	.405	.558	.573
Number of trip-level observations	533172	533172	533172	533172	352887	352887	352887	352887
Controls	No	Yes	Yes	Yes	No	Yes	Yes	Yes
Weather dummies	No	Yes	Yes	No	No	Yes	Yes	No
Driver fixed effects	No	No	Yes	Yes	No	No	Yes	Yes
Day fixed effects	No	No	No	Yes	No	No	No	Yes
Bus type × day fixed effects	No	No	No	Yes	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 until 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. The post-announcement dummy variable is one in the period from November 9, 2015, onwards (kickoff-event), zero otherwise. Drivers are considered to be in the post-feedback period when they have received at least one report in the past. For most drivers, this was at 15 December 2015 and after. The post-experimental period starts at November 15, 2016, when it was communicated to the treated drivers that they will no longer receive peer-comparison messages. The dependent variable fuel economy is measured in liters/100km and acceleration as the number of events per 10 kilometers. Standard errors are clustered by driver. Controls include: travel distance, route dummies, number of passengers and bus stops, and dummies for bus type, morning and evening rush hours, non-scheduled rides and having been coached. Weather dummies: daily temperature, wind and rainfall. 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). Standard errors in parentheses.

Table 7: In-Person Coaching Effects on Driving Performance

Outcome		Coefficient	p-values		
			Unadj.	Bonf.	Holm
Fuel economy	day of coaching	-0.6086	0.0000***	0.0000***	0.0000***
	day 1-7 after coaching	-0.3124	0.0000***	0.0000***	0.0000***
	post-coaching	-0.1235	0.0311**	0.3732	0.2799
Acceleration	day of coaching	-1.0803	0.000***	0.0000***	0.0000***
	day 1-7 after coaching	-0.6885	0.0000***	0.0000***	0.0000***
	post-coaching	-0.4469	0.0002***	0.0024***	0.0012***
Braking	day of coaching	-0.1094	0.0013***	0.0156**	0.0091***
	day 1-7 after coaching	-0.0244	0.3264	1.0000	1.0000
	post-coaching	-0.0159	0.5294	1.0000	1.0000
Cornering	day of coaching	-0.1901	0.0000***	0.0000***	0.0000***
	day 1-7 after coaching	-0.1043	0.0027***	0.0324**	0.0216**
	post-coaching	-0.0473	0.2118	1.0000	1.0000

Notes: Identification of in-person coaching effects on driving performance. The time period under consideration is the period for which we have complete logs available from all coaches (01/01/2015-30/04/2016). The dependent variable fuel economy is measured in liters/100km; acceleration, braking and cornering as the number of events per 10 kilometers. Reported are the DID effects of coaching on the day of coaching, in the first week following coaching and in the entire post-coaching period following coaching. Standard p-values as well as p-values that use a Bonferroni and Holm correction for multiple hypothesis testing are reported. Standard errors are clustered by driver. Full regression results are reported in Tables A11 and Table A15 in the appendix.
***(**,*) : the corresponding p-values are less than 1% (5% or 10%).

Table 8: Impact of Peer-Comparison Feedback, Conditional on Coaching

Dependent variable:	**Fuel Economy**		**Acceleration**		**Braking**		**Cornering**	
	Coached	Non-coached	Coached	Non-coached	Coached	Non-coached	Coached	Non-coached
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Post-announcement	-0.455***	-0.338***	-1.261***	-1.090***	-1.336***	-1.302***	-0.225***	-0.328***
	(0.072)	(0.068)	(0.168)	(0.175)	(0.048)	(0.050)	(0.051)	(0.059)
Post-feedback	0.171*	-0.023	-0.673***	-0.335	-0.159***	0.199***	-0.157**	-0.098
	(0.100)	(0.105)	(0.249)	(0.202)	(0.051)	(0.057)	(0.066)	(0.107)
Post-feedback×T2 [1n/0p]	0.117	-0.379**	0.260	-0.409	0.101	-0.130*	0.061	-0.047
	(0.135)	(0.155)	(0.356)	(0.318)	(0.071)	(0.073)	(0.111)	(0.116)
Post-feedback×T3 [1n/1p]	0.082	-0.117	0.183	-0.512	0.056	-0.088	0.075	0.081
	(0.107)	(0.181)	(0.284)	(0.417)	(0.068)	(0.106)	(0.090)	(0.119)
Post-feedback×T4 [3n/0p]	0.064	-0.304**	0.242	-1.012***	0.095	-0.282***	0.139*	-0.226
	(0.101)	(0.147)	(0.246)	(0.326)	(0.076)	(0.090)	(0.075)	(0.137)
Post-experiment	0.537***	0.476***	0.220	0.161	0.066	-0.003	-0.026	-0.092*
	(0.082)	(0.086)	(0.163)	(0.188)	(0.045)	(0.042)	(0.035)	(0.051)
R^2	0.498	0.502	0.561	0.519	0.233	0.219	0.379	0.407
# trip-level observations	232597	136482	164593	94667	164593	94667	167632	95938

Notes: The estimated regression specification includes the same controls, weather dummies and driver fixed effects as the specification estimated in columns (3) and (7) of Table 6. Full regression estimates reported in Tables A16-A19.
***(**,*) : statistically different from zero at the 1%-level (5%-level, 10%-level). Standard errors in parentheses.

Figures

Figure 1: Timeline Study

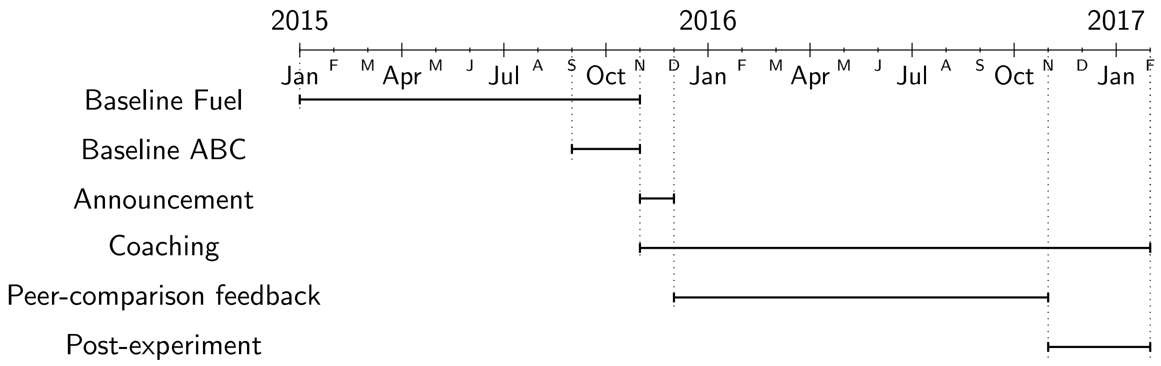
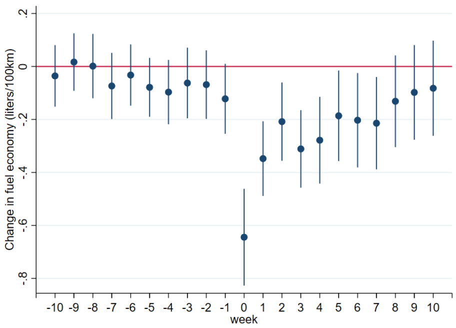
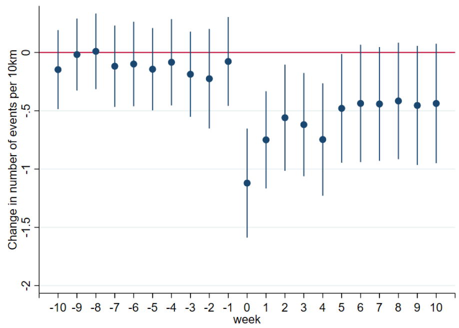


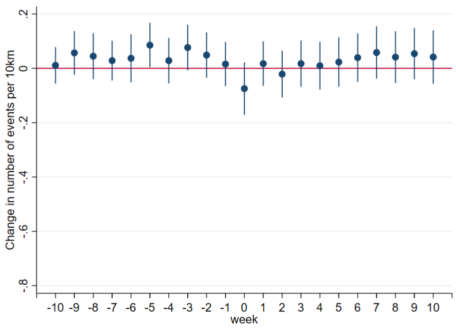
Figure 2: Temporal Effects In-Person Coaching



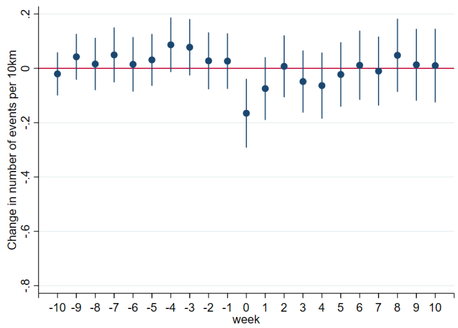
(a) Fuel economy



(b) Acceleration



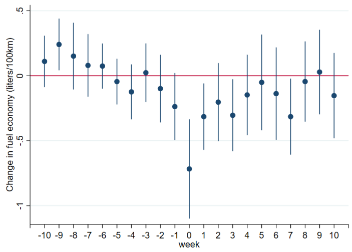
(c) Braking



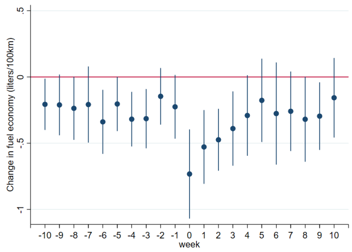
(d) Cornering

Notes: Driving performance in the 10 weeks before and after coaching based on trips with VDL and Intouro buses. The day of coaching itself is point 0 on the x-axis. The vertical spikes indicate 95% confidence intervals. The dependent variable fuel economy is measured in liters/100km and acceleration, braking and cornering as the number of events per 10 kilometers. Standard errors are clustered by driver. Controls include: travel distance, number of passengers and bus stops, dummies for non-scheduled rides, additional coaching sessions, bus types, morning and evening rush hours, and the interaction of bus type and day fixed effects. Coaches themselves are excluded from the sample.

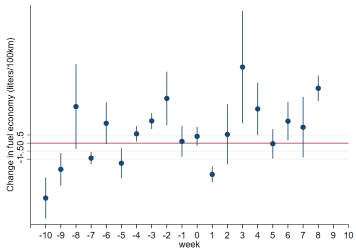
Figure 3: Temporal Effects In-Person Coaching at Coach Level: Fuel Economy



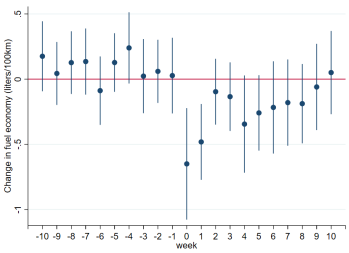
(a) Coach # 1 (n=37,935)



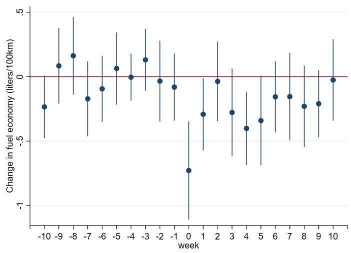
(b) Coach # 2 (n=39,217)



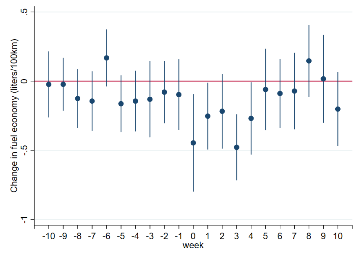
(c) Coach # 3 (n=1,633)



(d) Coach # 4 (n=37,376)

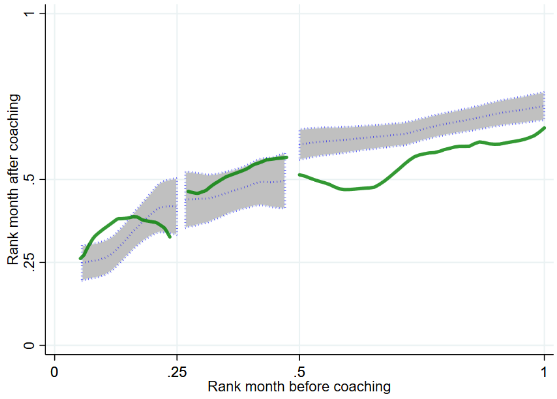


(e) Coach # 5 (n=27,345)

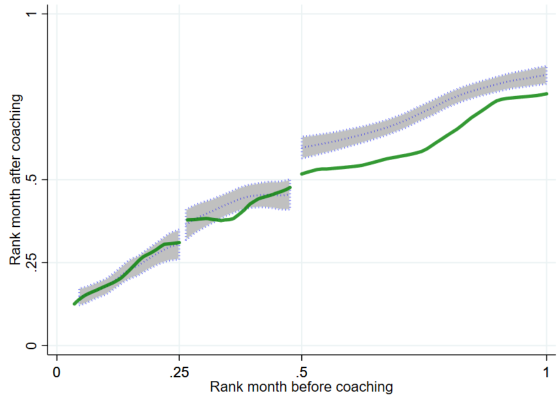


(f) Coach # 6 (n=42,174)

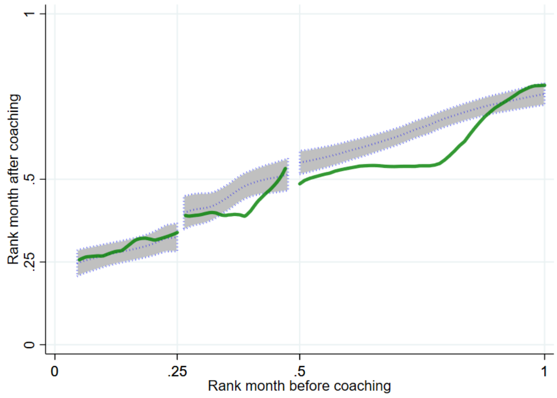
Figure 4: Differential Treatment Effects of Coaching



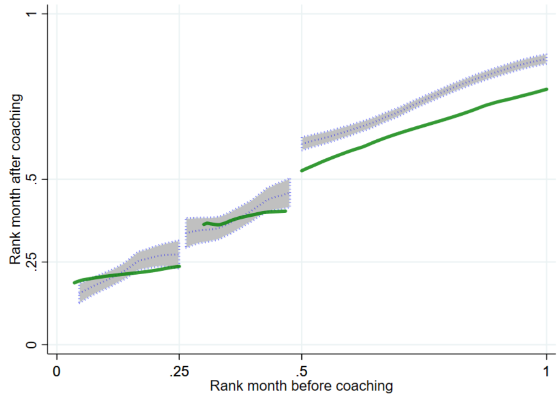
(a) Fuel economy



(b) Acceleration



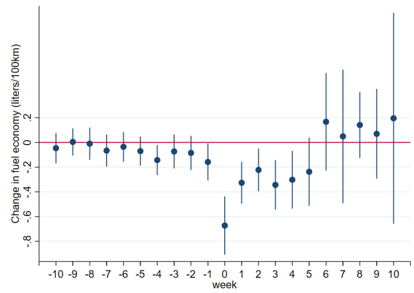
(c) Braking



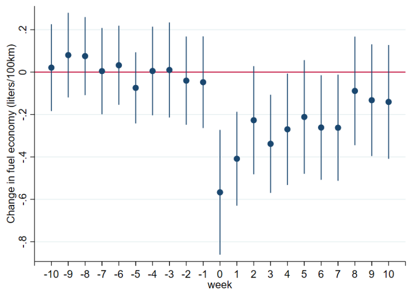
(d) Cornering

Note: Green (blue) : Average ranking of drivers (not) coached in month m in month $m-1$ (x-axis) and month $m+1$ (y-axis). The lines plot the non-parametric piece-wise local polynomial fit using an Epanechnikov kernel function. For non-coached drivers, the 95% confidence interval (shaded area) is shown as well.

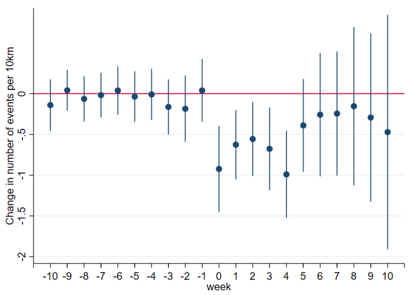
Figure 5: Impact In-Person Coaching With and Without Having First Received Written Feedback Reports



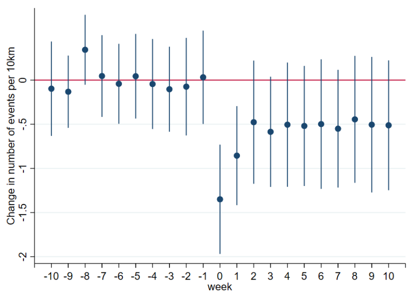
(a) Fuel economy: Coached before feedback



(b) Fuel economy: Coached after feedback

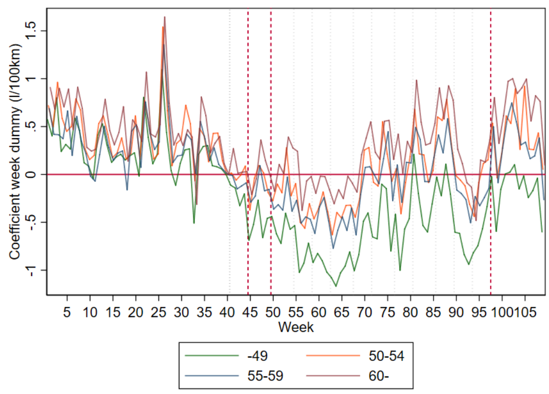


(c) Acceleration: Coached before feedback

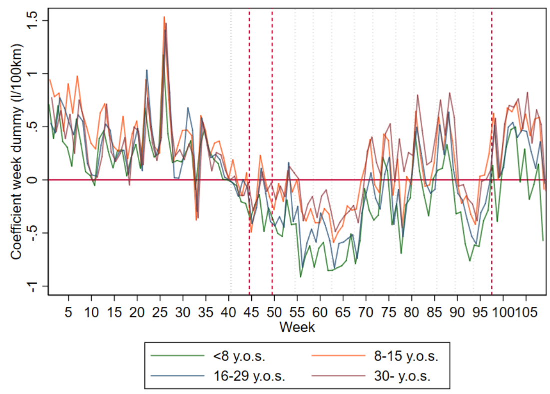


(d) Acceleration: Coached after feedback

Figure 6: Response in Fuel Economy to Written Feedback per Sub-Group of Drivers



(a) Drivers:. . . By age



(b) . . . By years of service

(c) . . . By gender

Notes: Coefficients week dummies on trips with VDL buses. The time period is from 01/01/2015 to 31/01/2017. The dashed lines indicate the launch date of the EcoManager program [09/11/2015], the distribution of the first feedback report [15/12/2015] and the distribution of the report with the final notification message [15/11/2016]. The dotted lines indicate the distribution of the intermediate monthly feedback reports. The first dotted line indicates the moment EcoManager promotion materials are send to the locations [05/10/2015]. Controls include: weather conditions, travel distance, route dummies, number of passengers and bus stops, having received coaching, and dummies for bus type, morning and evening rush hours and fill-in rides. A no-report indicator is included to capture drivers operating after 15/12/2015 (first feedback round) but who have not yet received their first report.

Online Appendix: Not for Publication

Improving Worker Productivity Through Tailored Performance Feedback:
Field Experimental Evidence from Bus Drivers

Gert-Jan Romensen and Adriaan R. Soetevent

A Construction Final Sample

The initial sample consists of 1,278,913 trip-level observations. Table A1 summarizes the steps taken to arrive at the final sample that is used in our analysis. The absence of a bus identifier has been the most important reason to discard observations (22.7% of the initial sample). Without such an identifier, it is not possible to link the trip to the outcome data of interest (fuel economy and ABC dimensions). Observations are also dropped when there are technical mismatches between the bus type and EOBR data (about 5%) or when there are between-trip inconsistencies at the driver level (3%). We also exclude extreme outcomes regarding punctuality or outcomes that suggest a temporary technical recording problem in the EOBR (< 0.5 percent). These comprise observations of fuel economy being less than one or more than eight (1,259 obs; 0.10%), a difference of more than one hour between actual and planned driving time (156 obs; 0.01%) and outcomes that are more than five standard deviations above the means of the ABC dimensions (4,003 obs; 0.31%).

Table A1: Cleaning Steps for Sample Construction

		% share of full sample
Full sample	**1,278,913**	
Reason for dropping observation:		
Duplicate observation (in terms of all variables)	(6,762)	0.50
No bus identifier	(290,737)	22.73
Bus type not eligible for EOBR	(34,870)	2.73
Error message from EOBR	(31,118)	2.46
Within-driver obs. with the same departure date/time	(37,575)	2.94
Very short rides (less than 1 kilometer)	(29,588)	2.31
Punctuality: more than 1 hour	(156)	0.01
Unreasonable outcomes of dependent variables:		
- Fuel eeconomy: less than 12.5 or more than 100	(1,259)	0.10
- ABC dimensions: more than 5 SDs above the mean	(4,003)	0.31
	842,845	

Notes: Fuel economy is measured in liters/100km. The ABC comfort dimensions are the number of events per 10 kilometers (fewer events mean better driving behavior). Trip punctuality is the difference between actual and planned driving time.

The company uses different bus types. For the trips in the final sample, Table A2 shows which bus type is used. The table reveals that the vast majority of trips (> 63%) is completed with a VDL bus. More specifically, over half of the trips are completed with the VDL 12 meter bus. The IRIS bus is used in about 30% of all trips. Most of these IRIS buses run on natural gas, which implies that no fuel economy score is recorded for these trips. These gas buses are almost exclusively used in the province's capital because of the lower CO_2 emissions of natural gas compared to diesel. A small minority (<10%) of trips is completed with a third bus type, the Mercedes Intouro.

Table A2: Fleet Information

Bus type	length (m)	fuel type	No. trips	% share of full sample
VDL AMBASSADOR ALE 106	10.6	diesel	76,815	9.11
VDL CITEA LLE 120	12.0	diesel	452,375	53.67
VDL CITEA XLE 145	14.5	diesel	5,830	0.69
IRISBUS CITELIS 10,5 M	10.5	diesel	22,882	2.71
IRISBUS CITELIS 10,5 M CNG	10.5	natural gas	70,341	8.35
IRISBUS CITELIS 12 M	12.0	diesel	41,048	4.87
IRISBUS CITELIS 12 M CNG	12.0	natural gas	113,731	13.49
MERCEDES BENZ INTOURO	13.0	diesel	59,823	7.10
Analysis set			**842,845**	
VDL			535,020	63.5
IRISBUS			248,002	29.4
INTOURO			59,823	7.1

Table A3 shows per bus type when the start of the recording of data. Data on fuel economy are available for a somewhat longer time period than the outcomes on the ABC-comfort dimensions.

Table A3: Start Date of Data Recording per Bus Type

	VDL	Intouro	IRIS
Fuel consumption	1 Jan. 2015	1 Jan. 2015	n.a.
Acceleration	1 Sep. 2015	9 Nov. 2015	1 Sep. 2015
Braking	1 Sep. 2015	9 Nov. 2015	1 Sep. 2015
Cornering	1 Sep. 2015	1 Sep. 2015	9 Nov. 2015
No. of trips	535,020	59,823	248,002

Notes: The number of trips for which the ABC events are recorded is lower than the total number of trips mentioned in the table, because recording of ABC events commenced only in Sept. 2015. This explains the difference between the number of observations mentioned in this table and those in the regression tables of Section 5.

A.1 Intermediate Changes in ABC settings

During the period of data collection, the company changed some of the threshold settings for the ABC dimensions. An increase (decrease) in the threshold has the effect of reducing (increasing) the number of recorded events. The company has provided a detailed list of when which threshold has been changed on which bus type(s). This list is presented in Table A4.

We note that the VDL buses did not have any changes in settings throughout the period of data collection, except for the braking threshold, which was increased on Oct. 16, 2015 and subsequently slightly decreased on Nov. 5. Especially the Oct. 16 change seems to result in a drop of the number of recorded events, as can be seen in Figure A1b, which shows by bus type the development of the scores in ABC dimensions and fuel economy (weekly averages).

As Figure A1a-c (right axis) show, the IRIS buses record an importantly larger number of events in all three dimensions than the other two bus types. This is due to the fact that this is the bus type of choice in the province's capital, where the routes are characterized by many road bends and stops. Unfortunately, both the acceleration and cornering thresholds for the IRIS buses were increased at Dec. 11, just around the time the drivers received their first report. We therefore cannot identify whether the drop we observe in Figures A1b and c for the number of braking and cornering events is due to the report or the change in settings. For cornering, no data are available for the period prior to the kickoff event. Due to these issues, we exclude the IRIS bus from the analysis in the main text.

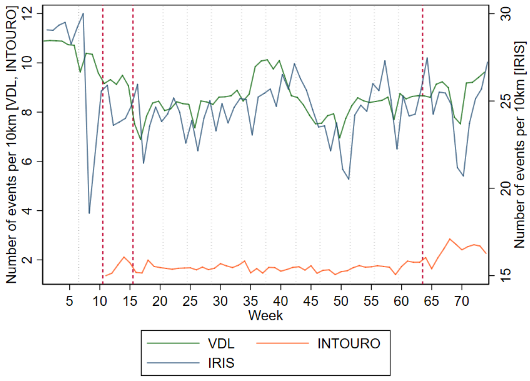
The Intouro buses experienced a number of recalibrations, but all before the date the first feedback report was received by the drivers and mostly comprised only two to four buses out of a total of 29 Intouro buses. Because calibration for these buses was late, we have no records for acceleration and braking for the period prior to the kickoff event, as Figures A1a and b clearly show.

Other than the ABC dimensions, the fuel economy records of both the VDL and Intouro buses cover the entire period from January 2015 till January 2017 and have not been subject to any change in measurement. For completeness, Figures A1d shows the weekly average fuel economy for both the VDL and Intouro buses.

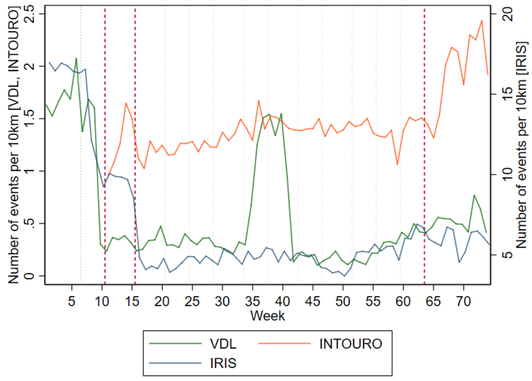
Table A4: Change in Threshold Settings

Bus type	date of change	dimension affected	nature of threshold change
IRISBUS 12 M	Dec. 11, 2015	A	decrease
IRISBUS 12 M CNG	Dec. 11, 2015	A	decrease
IRISBUS 10,5 M	Dec. 11, 2015	C	increase
IRISBUS 10,5 M CNG	Dec. 11, 2015	C	increase
VDL AMBASSADOR ALE 106	Oct. 16, 2015	B	increase
	Nov. 5, 2015	B	decrease
VDL CITEA LLE 120	Oct. 16, 2015	B	increase
MERCEDES BENZ INTOURO			
bus no. 7503, 7504	Sept. 16, 2015	ABC	recalibration
bus no. 7503, 7504	Oct. 02, 2015	ABC	recalibration
bus no. 7501, 7502, 7503	Nov. 25, 2015	ABC	recalibration
all	Dec. 03, 2015	AB	recalibration

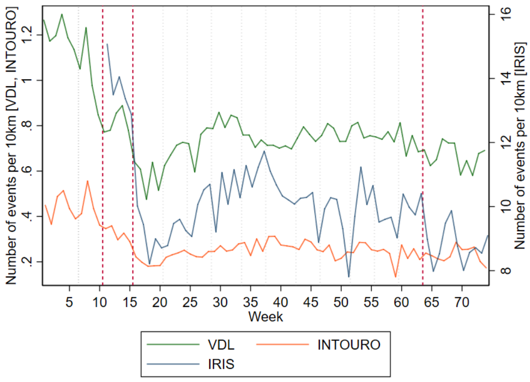
Figure A1: Development Over Time in ABC Dimensions (No. Events) and Fuel Economy (liters/100km) – Weekly Averages by Bus Type



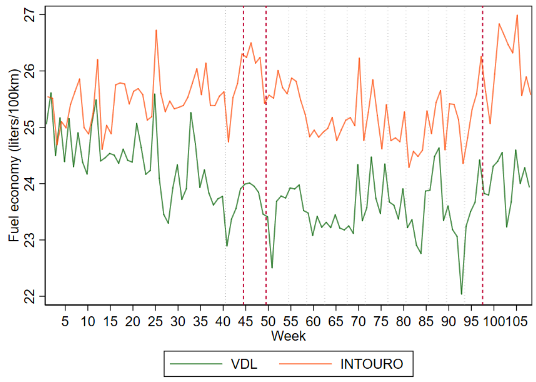
(a) Acceleration



(b) Braking



(c) Cornering



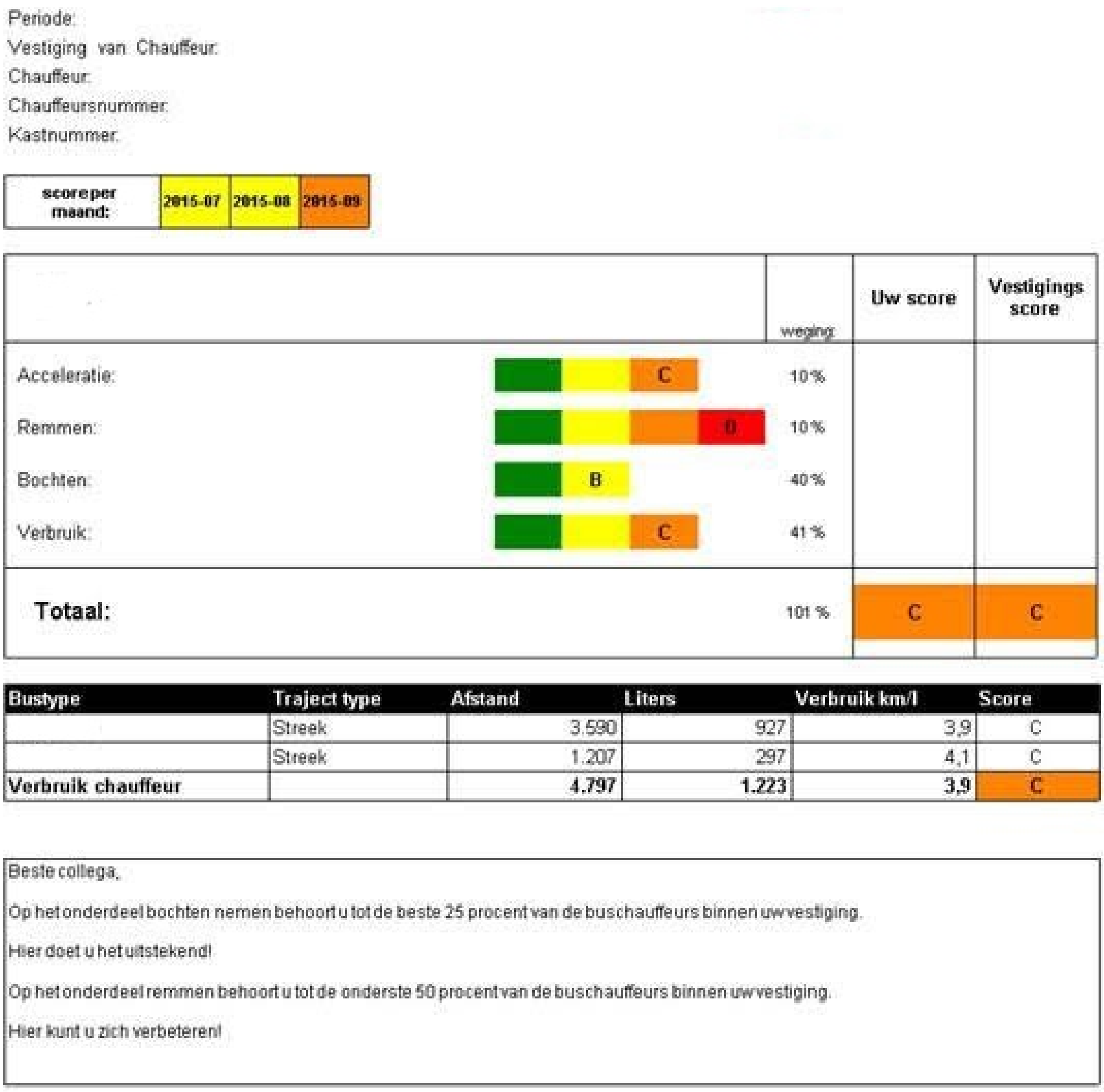
(d) Fuel economy

Note: Averages are calculated based on all trips in the analysis set. The red dashed lines indicate the launch date of the EcoManager program [09/11/2015], the distribution of the first feedback report [15/12/2015] and the distribution of the report with the final notification message [15/11/2016]. The dotted lines indicate the distribution of the intermediate monthly feedback reports. The first dotted line indicates the moment EcoManager promotion materials are send to the locations [05/10/2015].

B Sample Feedback Report

Figure A2 reproduces a specimen of the feedback report drivers received once a month between December 2015 - November 2016.

Figure A2: Sample Feedback Report



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

scoreper 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!

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

C Results on Braking and Cornering

Table A5: Targeted Peer-Comparison Feedback Effects on Driving Performance: Braking and Cornering

Dependent variable:	**Braking**				**Cornering**			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Post-announcement	-1.267***	-1.268***	-1.292***		-0.334***	-0.272***	-0.317***	
	(0.027)	(0.027)	(0.027)		(0.029)	(0.033)	(0.028)	
T2 (1n/0p)	-0.081	-0.084			-0.075	-0.080		
	(0.059)	(0.058)			(0.117)	(0.116)		
T3 (1n/1p)	-0.083	-0.082			-0.152	-0.158		
	(0.063)	(0.063)			(0.121)	(0.117)		
T4 (3n/0p)	-0.029	-0.030			0.012	0.007		
	(0.065)	(0.063)			(0.121)	(0.120)		
Post-feedback	0.070**	0.020	0.000		-0.145***	-0.160***	-0.193***	
	(0.033)	(0.035)	(0.036)		(0.052)	(0.058)	(0.051)	
Post-feedback × T2 (1n/0p)	0.063	0.071	0.052	0.047	0.044	0.048	0.063	0.059
	(0.050)	(0.049)	(0.051)	(0.049)	(0.070)	(0.070)	(0.068)	(0.068)
Post-feedback × T3 (1n/1p)	0.003	0.015	0.009	0.015	0.114*	0.118*	0.110*	0.108*
	(0.054)	(0.052)	(0.053)	(0.051)	(0.069)	(0.067)	(0.065)	(0.064)
Post-feedback × T4 (3n/0p)	-0.017	0.001	-0.020	-0.014	0.035	0.040	0.063	0.062
	(0.053)	(0.052)	(0.051)	(0.050)	(0.068)	(0.068)	(0.063)	(0.062)
Post-experiment	0.033	-0.003	-0.011		-0.100***	-0.106***	-0.051**	
	(0.028)	(0.027)	(0.027)		(0.034)	(0.036)	(0.025)	
Post-experiment × T2 (1n/0p)	0.005	-0.004	0.009	-0.001	0.067	0.064	0.004	0.005
	(0.040)	(0.038)	(0.036)	(0.036)	(0.043)	(0.042)	(0.033)	(0.033)
Post-experiment × T3 (1n/1p)	0.099**	0.107***	0.099***	0.089***	0.034	0.043	-0.026	-0.029
	(0.039)	(0.035)	(0.035)	(0.034)	(0.042)	(0.040)	(0.032)	(0.031)
Post-experiment × T4 (3n/0p)	0.080**	0.059	0.058	0.041	0.068	0.076	-0.007	-0.004
	(0.039)	(0.039)	(0.039)	(0.038)	(0.052)	(0.050)	(0.040)	(0.040)
Constant	1.751***	1.367***	2.313***	2.157***	1.142***	1.302***	1.637***	1.735***
	(0.050)	(0.055)	(0.059)	(0.085)	(0.090)	(0.096)	(0.056)	(0.078)
R^2	.0841	.2	.224	.28	.0264	.0912	.388	.397
Number of trip-level observations	352887	352887	352887	352887	359066	359066	359066	359066
Controls	No	Yes	Yes	Yes	No	Yes	Yes	Yes
Weather dummies	No	Yes	Yes	No	No	Yes	Yes	No
Driver fixed effects	No	No	Yes	Yes	No	No	Yes	Yes
Day fixed effects	No	No	No	Yes	No	No	No	Yes
Bus type × day fixed effects	No	No	No	Yes	No	No	No	Yes

Notes: Identification of the treatment effects on driving performance. The time period under consideration is from 01/01/2015, until 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. The post-announcement dummy variable is equal to one in the period from 09/11/2015 onwards, zero otherwise. Drivers are considered to be in the post-feedback period when they have received at least one report in the past. For most drivers, this was at 15/12/2015 and after. The post-experimental period starts at 15/11/2016 when it was communicated to the treated drivers that they will no longer receive peer-comparison messages. The dependent variables braking and cornering are measured as the number of events per 10 kilometers. Standard errors are clustered by driver. Controls include: travel distance, route dummies, number of passengers and bus stops, and dummies for bus type, morning and evening rush hours, fill-in rides and having been coached. Weather dummies: daily temperature, wind and rainfall. A no-report indicator is included to capture drivers operating after 15/12/2015 (first feedback round) but who have not yet received their first report.

***(**,*) : statistically different from zero at the 1%-level (5%-level, 10%-level). Standard errors in parentheses.

D Additional Tables and Figures

D.1 Model of the estimation of the residual standard deviation

Let the outcome variable of interest, Y_{it} (fuel economy or ABC), indexed by driver (i) and trip (t), be given by:

$$Y_{it} \quad = \quad X_{it} \cdot \beta + \mu_i + f(s_{it}) + g(z_{it}) + \epsilon_{it}. \tag{3}$$

In this specification, X_{it} includes all observable determinants of driver performance, day, driver and route fixed effect. The μ_i parameters denote driver fixed effects that absorb all time-invariant unobservables at the driver level and the functions $f()$ and $g()$ reflect the potential impact of having been exposed to the peer-comparison treatment (s_{it}) and coaching (z_{it}), respectively. Especially for the ABC-dimensions our baseline period before treatment is too small for a reliable estimation of the residual variation. Hence, similar to Chetty, Friedman and Rockoff (2014), we use the entire sample to compute driver's residual outcomes. The residual r_{it} in outcomes after accounting for the observable determinants of driver performance, day, driver and route fixed effect is constructed as follows:

$$r_{it} \quad = \quad Y_{it} - X_{it} \cdot \beta = \mu_i + f(s_{it}) + g(z_{it}) + \epsilon_{it}. \tag{4}$$

The standard deviation of the residual variation is computed as:

$$\sigma_r = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(r_i - \bar{r})^2},$$

with $r_i = \frac{1}{n_i}\sum_t^{n_i} r_{it}$ driver i's mean residual over n_t trips and $\bar{r} = \frac{1}{n}\sum_t^{n} r_i$ the mean residual across drivers. In case the impact of treatment (written feedback or in-person coaching) is correlated with unobservables at the driver level, σ_r will over- or understate the residual variation across drivers. When for example coaches tend to select on average worse drivers (for which we do not find evidence) or when worse drivers benefit more from coaching (for which we do find some evidence), σ_r will be a conservative, downward biased estimate of the actual across driver variation in μ_i's.

D.2 Determinants of ABC Outcomes

Table A6: Determinants of Acceleration

Dependent variable:	Acceleration				
	(1)	(2)	(3)	(4)	(5)
VDL 10m	0.953***	0.969***	1.090***	1.139***	1.086***
	(0.073)	(0.073)	(0.071)	(0.072)	(0.048)
VDL 14m	-0.080	-0.127	-0.189	-0.301	-0.172
	(0.149)	(0.146)	(0.157)	(0.191)	(0.198)
Intouro	-1.454***	-1.414***	-1.756***	-1.472***	-1.235***
	(0.210)	(0.213)	(0.205)	(0.214)	(0.224)
Rush hour 7-10am			-0.896***	-0.909***	-0.844***
			(0.144)	(0.146)	(0.127)
Rush hour 4-7pm			1.182***	1.199***	1.146***
			(0.157)	(0.156)	(0.111)
Non-scheduled trip			0.353*	0.461**	0.624***
			(0.202)	(0.208)	(0.165)
No. of stops per km.			2.274***	2.376***	2.361***
			(0.194)	(0.188)	(0.169)
Urban trip			0.000	0.000	0.000
			(.)	(.)	(.)
Trip length (in km.)			-0.025***	-0.021***	-0.021***
			(0.002)	(0.002)	(0.002)
Ln(No. of passengers)			1.858***	1.728***	1.648***
			(0.081)	(0.080)	(0.072)
Punctuality			-0.039***	-0.031**	0.049***
			(0.015)	(0.015)	(0.007)
Constant	13.267***	14.002***	6.341***	4.694***	4.718***
	(0.188)	(0.199)	(0.457)	(0.429)	(0.357)
R^2	.514	.516	.537	.543	.602
Number of trip-level observations	513866	513866	513866	513866	513866
Weather dummies	No	Yes	Yes	No	No
Driver fixed effects	No	No	No	No	Yes
Day fixed effects	No	No	No	Yes	Yes
Route fixed effects	No	No	Yes	Yes	Yes

Notes: Dependent variable: Number of acceleration events per 10 km. Default categories: Scheduled trips in rural areas outside rush hours completed with a VDL 12m bus. Standard errors are clustered by driver.
***(**,*) : statistically different from zero at the 1%-level (5%-level, 10%-level). Standard errors in parentheses.

Table A7: Determinants of Braking

Dependent variable:	Braking				
	(1)	(2)	(3)	(4)	(5)
VDL 10m	0.959***	0.969***	0.984***	1.037***	1.017***
	(0.032)	(0.032)	(0.032)	(0.032)	(0.031)
VDL 14m	0.232***	0.195***	0.178***	0.145***	0.197***
	(0.029)	(0.033)	(0.037)	(0.055)	(0.071)
Intouro	3.758***	3.776***	3.632***	3.741***	3.771***
	(0.240)	(0.240)	(0.239)	(0.236)	(0.227)
Rush hour 7-10am			0.005	-0.008	-0.024
			(0.054)	(0.054)	(0.033)
Rush hour 4-7pm			0.207***	0.221***	0.252***
			(0.077)	(0.076)	(0.046)
Non-scheduled trip			-0.061	-0.053	-0.078
			(0.080)	(0.077)	(0.085)
No. of stops per km.			0.275**	0.324***	0.321***
			(0.118)	(0.113)	(0.089)
Urban trip			0.000	0.000	0.000
			(.)	(.)	(.)
Trip length (in km.)			-0.005***	-0.002***	-0.002***
			(0.001)	(0.001)	(0.000)
Ln(No. of passengers)			0.384***	0.287***	0.280***
			(0.037)	(0.034)	(0.031)
Punctuality			-0.030***	-0.025***	-0.000
			(0.006)	(0.006)	(0.002)
Constant	0.956***	1.303***	-0.001	0.121	0.115
	(0.066)	(0.076)	(0.253)	(0.220)	(0.185)
R^2	.463	.467	.473	.492	.575
Number of trip-level observations	513866	513866	513866	513866	513866
Weather dummies	No	Yes	Yes	No	No
Driver fixed effects	No	No	No	No	Yes
Day fixed effects	No	No	No	Yes	Yes
Route fixed effects	No	No	Yes	Yes	Yes

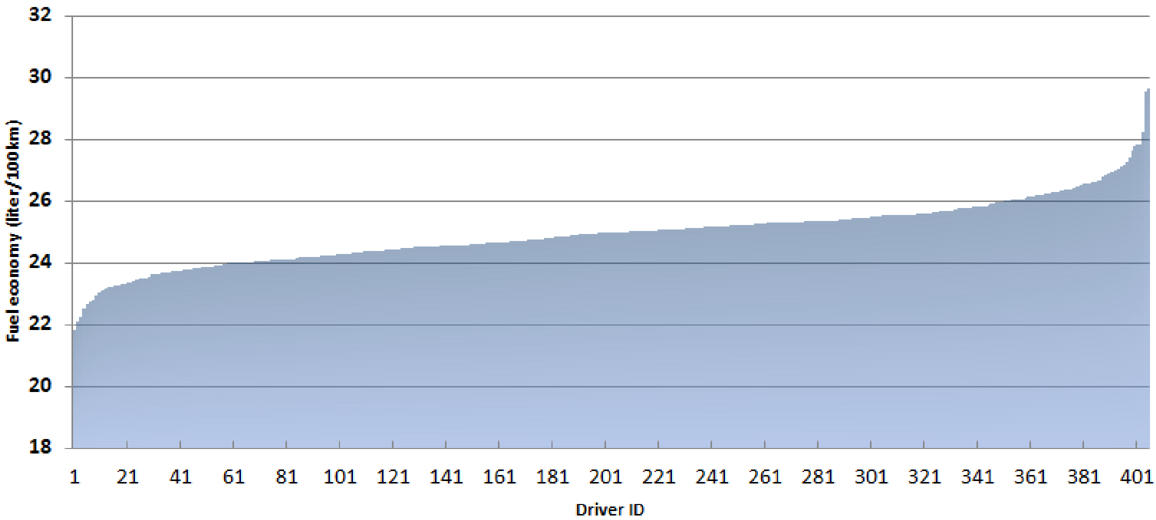
Notes: Dependent variable: Number of braking events per 10 km. Default categories: Scheduled trips in rural areas outside rush hours completed with a VDL 12m bus. Standard errors are clustered by driver.
***(**,*) : statistically different from zero at the 1%-level (5%-level, 10%-level). Standard errors in parentheses.

Table A8: Determinants of Cornering

Dependent variable:	Cornering				
	(1)	(2)	(3)	(4)	(5)
VDL 10m	-0.085***	-0.079***	-0.110***	-0.078***	-0.064***
	(0.018)	(0.018)	(0.018)	(0.019)	(0.013)
VDL 14m	-0.164***	-0.132***	-0.109**	-0.125	-0.103
	(0.039)	(0.043)	(0.051)	(0.077)	(0.117)
Intouro	-0.343***	-0.320***	-0.463***	-0.371***	-0.279***
	(0.043)	(0.044)	(0.059)	(0.065)	(0.055)
Rush hour 7-10am			-0.130	-0.139	-0.172***
			(0.101)	(0.099)	(0.049)
Rush hour 4-7pm			0.164	0.171	0.177***
			(0.116)	(0.115)	(0.045)
Non-scheduled trip			-0.116	-0.108	-0.327**
			(0.119)	(0.117)	(0.159)
No. of stops per km.			-0.509***	-0.570***	-0.518***
			(0.134)	(0.134)	(0.095)
Urban trip			0.000	0.000	0.000
			(.)	(.)	(.)
Trip length (in km.)			-0.004***	-0.002***	-0.003***
			(0.001)	(0.001)	(0.000)
Ln(No. of passengers)			0.151***	0.098***	0.094***
			(0.031)	(0.032)	(0.023)
Punctuality			-0.121***	-0.119***	-0.055***
			(0.014)	(0.014)	(0.006)
Constant	3.096***	3.386***	3.797***	1.610***	1.467***
	(0.163)	(0.179)	(0.318)	(0.244)	(0.177)
R^2	.5	.502	.506	.513	.676
Number of trip-level observations	513866	513866	513866	513866	513866
Weather dummies	No	Yes	Yes	No	No
Driver fixed effects	No	No	No	No	Yes
Day fixed effects	No	No	No	Yes	Yes
Route fixed effects	No	No	Yes	Yes	Yes

Notes: Dependent variable: Number of cornering events per 10 km. Default categories: Scheduled trips in rural areas outside rush hours completed with a VDL 12m bus. Standard errors are clustered by driver.
***(**,*) : statistically different from zero at the 1%-level (5%-level, 10%-level). Standard errors in parentheses.

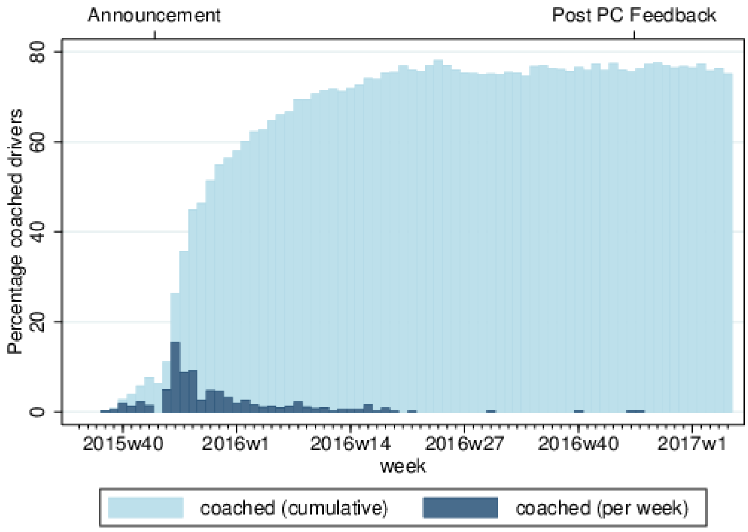
Figure A3: Fuel economy: Estimates Driver Fixed Effects



Note: Estimates based on trips completed with VDL and Intouro buses.

D.3 Timing of Coaching Sessions

Figure A4: Time of First Coaching



Notes: Moment of first coaching for drivers. Dark blue bars indicate the drivers who received their first coaching during a specific week as a share of the total number of drivers operating during that week. The light blue bars depict the cumulative share of coached drivers operating during a week. Feedback was announced Nov. 9, 2015 and first distributed as a monthly report after Dec. 15, 2015. Peer-comparison messages were removed from the reports from Nov. 2016 onwards.

Two things in Figure A4 related to the coaching program deserve some further explanation. First, the cumulative share of coached drivers operating during a week is more or less flat after April 2016. We have complete coach logs for the period till 30 April, 2016. Some coaches indicated that they no longer provided or kept track of coaching after April 2016. In our evaluation of the coaching program, we therefore restrict attention to the period until 30 April 2016. Second, 30 drivers (10% of all coached drivers) received coaching prior to the feedback announcement.

D.4 Driver Exposure to Targeted Feedback

The tailored nature of the messages is illustrated in Table A9. 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).[43] It highlights

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

the flexible design of the treatments. Each treated driver is assigned an individualized message combination which points to behaviors that require attention. In treatment 1 (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 A5. 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.

Table A9: Incidence and Composition of Targeted Peer-Comparison Messages

	T1 (1n/0p)		**T2 (1n/1p)**				**T3 (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)

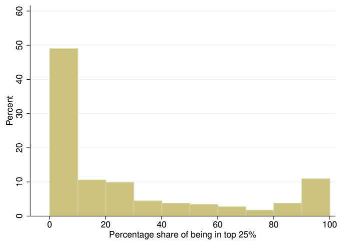
Notes: Panel A reports the percentage share of drivers receiving one of the possible message combinations in each treatment (conditional on receiving a feedback report). Drivers do not receive a report when they were absent in the previous month (on which the report is based). Panel B shows the composition 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. 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.

Table A10: Change in ranking month $(m+1)$ vs. m following relative performance feedback on performance indicator in month m.

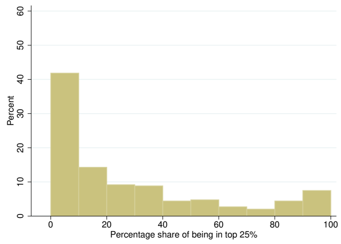
Dep. var		Rank month $(m+1)$			
		Acceleration			
	Message	0.001	-0.007	-0.086**	0.054**
		(0.02)	(0.016)	(0.036)	(0.022)
	Rank month m	0.725***	0.644***	0.851***	1.105***
		(0.079)	(0.053)	(0.075)	(0.22)
	obs.	464	521	487	197
	R^2	0.2503	0.2195	0.2884	0.214
Sample	Rank month m:	bottom-50%	bottom-50%	bottom-50%	top-25%
	Treatment:	T2[0p1n]	T3[1p1n]	T4[0p3n]	T3[1p1n]
		Braking			
	Message	-0.043*	-0.044**	-0.063	0.113***
		(0.022)	(0.019)	(0.048)	(0.036)
	Rank month m	0.704***	0.723***	0.712***	0.581**
		(0.096)	(0.062)	(0.113)	(0.253)
	obs.	468	516	492	204
	R^2	0.1928	0.216	0.1635	0.1248
Sample	Rank month m:	bottom-50%	bottom-50%	bottom-50%	top-25%
	Treatment:	T2[0p1n]	T3[1p1n]	T4[0p3n]	T3[1p1n]
		Cornering			
	Message	-0.003	-0.008	-0.014	0.012
		(0.012)	(0.012)	(0.031)	(0.019)
	Rank month m	0.911***	0.911***	0.808***	0.907***
		(0.043)	(0.049)	(0.062)	(0.148)
	obs.	470	508	500	196
	R^2	0.5877	0.524	0.4244	0.2048
Sample	Rank month m:	bottom-50%	bottom-50%	bottom-50%	top-25%
	Treatment:	T2[0p1n]	T3[1p1n]	T4[0p3n]	T3[1p1n]

Notes: All regressions include a constant. Standard errors clustered by driver.
***(**,*) : statistically different from zero at the 1%-level (5%-level, 10%-level). Standard errors in parentheses.

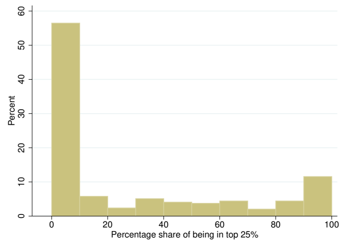
Figure A5: Share of Feedback Rounds in Top 25% or Bottom 50% for Treated Drivers



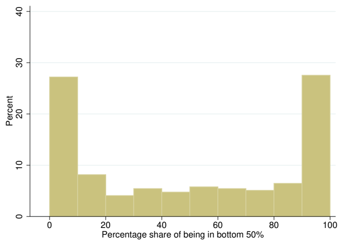
(a) Top 25%: acceleration



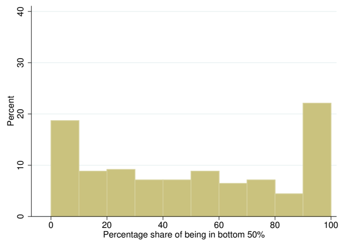
(b) Top 25%: braking



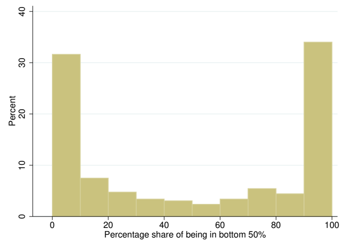
(c) Top 25%: cornering



(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.

E Further Results: In-Person Coaching

Table A11: In-Person Coaching Effects on Driving Performance: Fuel Economy and Acceleration

Dependent variable:	Fuel Economy			Acceleration		
	(1)	(2)	(3)	(4)	(5)	(6)
Post-coaching	-0.168***	-0.124**		-0.468***	-0.447***	
	(0.064)	(0.057)		(0.134)	(0.120)	
Day of first coaching			-0.609***			-1.080***
			(0.085)			(0.165)
1 –7 days after			-0.312***			-0.689***
			(0.058)			(0.120)
8 –14 days after			-0.174***			-0.467***
			(0.064)			(0.137)
15 –21 days after			-0.278***			-0.513***
			(0.063)			(0.133)
22 –28 days after			-0.246***			-0.654***
			(0.073)			(0.149)
29 –35 days after			-0.154**			-0.371***
			(0.077)			(0.138)
36 –42 days after			-0.171**			-0.335**
			(0.081)			(0.156)
43 –49 days after			-0.182**			-0.349**
			(0.079)			(0.159)
50 –56 days after			-0.099			-0.364**
			(0.077)			(0.172)
57 –63 days after			-0.066			-0.406**
			(0.081)			(0.166)
64 –70 days after			-0.050			-0.336*
			(0.083)			(0.185)
$>$ 70			0.048			-0.134
			(0.076)			(0.180)
Number of trip-level observations	352253	352253	352253	187127	187127	187127
Controls	No	Yes	Yes	No	Yes	Yes
Driver fixed effects	No	Yes	Yes	No	Yes	Yes
Day fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Bus type $\times$ day fixed effects	No	Yes	Yes	No	Yes	Yes

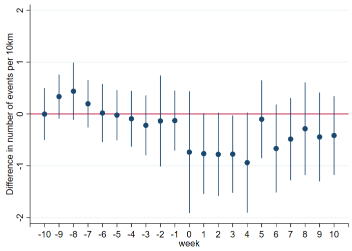
Notes: Identification of in-person coaching effects on driving performance. The time period under consideration is the period for which we have complete logs available from all coaches (01/01/2015-30/04/2016). Post-coaching identifies the day of first coaching and the full period thereafter for each coached driver. The dependent variable fuel economy is measured in liters/100km and acceleration as the number of events per 10 kilometers. Standard errors are clustered by driver. Controls include: travel distance, route dummies, number of passengers and bus stops, and dummies for bus type, morning and evening rush hours, non-scheduled rides and having been coached.
***(**,*) : statistically different from zero at the 1%-level (5%-level, 10%-level). Standard errors in parentheses.

Table A12: In-Person Coaching Effects on Driving Performance: Braking and Cornering

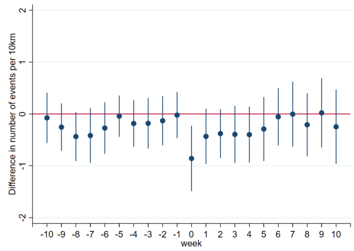
Dependent variable:	**Braking**			**Cornering**		
	(1)	(2)	(3)	(4)	(5)	(6)
Post-coaching	-0.009	-0.016		-0.047	-0.047	
	(0.026)	(0.025)		(0.038)	(0.038)	
Day of first coaching			-0.109***			-0.190***
			(0.034)			(0.039)
1 –7 days after			-0.024			-0.104***
			(0.025)			(0.035)
8 –14 days after			-0.063**			-0.027
			(0.027)			(0.035)
15 –21 days after			-0.016			-0.079**
			(0.025)			(0.037)
22 –28 days after			-0.031			-0.098**
			(0.026)			(0.040)
29 –35 days after			-0.018			-0.054
			(0.028)			(0.038)
36 –42 days after			-0.004			-0.020
			(0.029)			(0.045)
43 –49 days after			0.017			-0.039
			(0.031)			(0.047)
50 –56 days after			-0.007			0.012
			(0.031)			(0.051)
57 –63 days after			0.009			-0.030
			(0.030)			(0.051)
64 –70 days after			-0.005			-0.015
			(0.032)			(0.052)
> 70			0.020			0.018
			(0.037)			(0.055)
Number of trip-level observations	187127	187127	187127	187127	187127	187127
Controls	No	Yes	Yes	No	Yes	Yes
Driver fixed effects	No	Yes	Yes	No	Yes	Yes
Day fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Bus type $\times$ day fixed effects	No	Yes	Yes	No	Yes	Yes

Notes: Identification of in-person coaching effects on driving performance. The time period under consideration is the period for which we have complete logs available from all coaches (01/01/2015-30/04/2016). Post-coaching identifies the day of first coaching and the full period thereafter for each coached driver. The dependent variables braking and cornering are measured as the number of events per 10 kilometers. Standard errors are clustered by driver. Controls include: travel distance, route dummies, number of passengers and bus stops, and dummies for bus type, morning and evening rush hours, fill-in rides and having been coached. Weather dummies: daily temperature, wind and rainfall. ***(**,*) : statistically different from zero at the 1%-level (5%-level, 10%-level). Standard errors in parentheses.

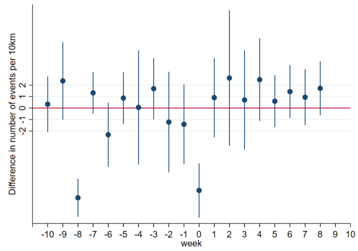
Figure A6: Temporal Effects In-Person Coaching at Coach Level: Acceleration



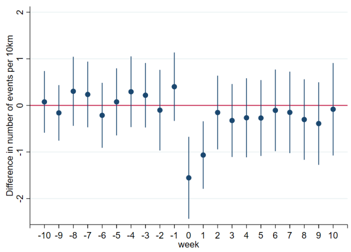
(a) Coach # 1



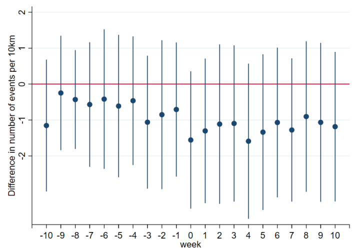
(b) Coach # 2



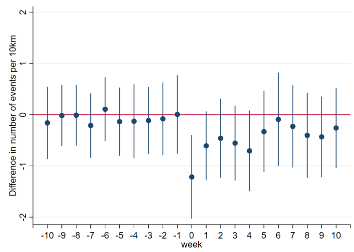
(c) Coach # 3



(d) Coach # 4



(e) Coach # 5



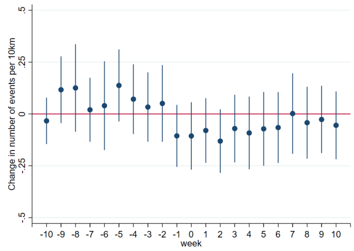
(f) Coach # 6

Table A13: Change in ranking month $(m+1)$ vs. $(m-1)$ following coaching in month m.

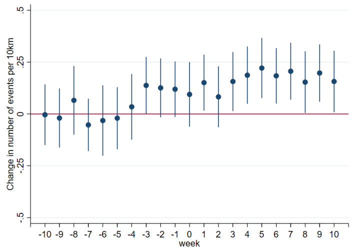
Dep. var	Rank month $(m+1)$			
		Acceleration		
	Coached month m	-0.006	-0.012	-0.074**
		(0.031)	(0.035)	(0.029)
	Rank month $(m-1)$	1.195***	0.697***	0.653***
		(0.11)	(0.122)	(0.043)
	obs.	754	736	1744
	R^2	0.1765	0.0495	0.2018
Sample	Rank month $(m-1)$:	top-25%	25-50%	bottom-50%
		Braking		
	Coached month m	0.004	-0.009	-0.068**
		(0.049)	(0.039)	(0.029)
	Rank month $(m-1)$	0.694***	0.527***	0.64***
		(0.141)	(0.136)	(0.047)
	obs.	751	731	1752
	R^2	0.0392	0.0229	0.1576
Sample	Rank month (m-1):	top-25%	25-50%	bottom-50%
		Cornering		
	Coached month m	-0.004	-0.024	-0.078***
		(0.021)	(0.042)	(0.021)
	Rank month $(m-1)$	0.845***	0.758***	0.843***
		(0.091)	(0.094)	(0.031)
	obs.	740	739	1755
	R^2	0.1478	0.0797	0.4318
Sample	Rank month $(m-1)$:	top-25%	25-50%	bottom-50%
		Fuel economy		
	Coached month m	0.007	0.074	-0.087*
		(0.088)	(0.055)	(0.046)
	Rank month $(m-1)$	0.931***	0.451**	0.565***
		(0.196)	(0.197)	(0.062)
	obs.	351	343	833
	R^2	0.0607	0.0184	0.1304
Sample	Rank month $(m-1)$:	top-25%	25-50%	bottom-50%

Notes: All regressions include a constant. Standard errors clustered by driver.
***(**,*) : statistically different from zero at the 1%-level (5%-level, 10%-level). Standard errors in parentheses.

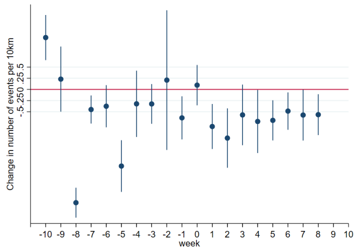
Figure A7: Temporal Effects In-Person Coaching at Coach Level: Braking



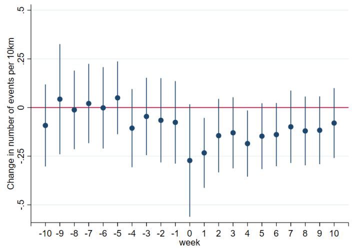
(a) Coach # 1



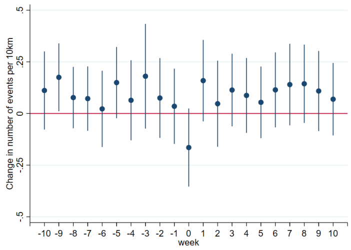
(b) Coach # 2



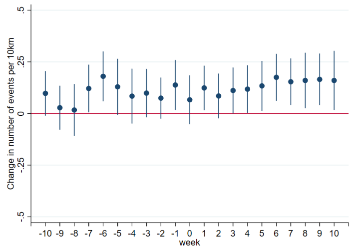
(c) Coach # 3



(d) Coach # 4

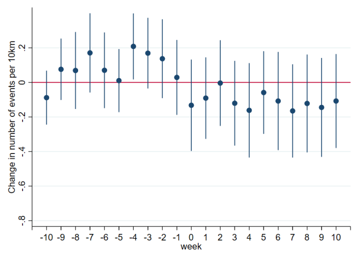


(e) Coach # 5

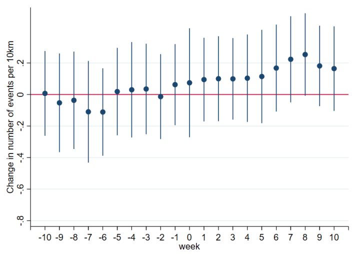


(f) Coach # 6

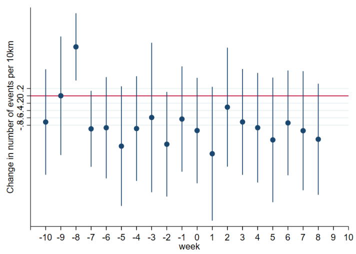
Figure A8: Temporal Effects In-Person Coaching at Coach Level: Cornering



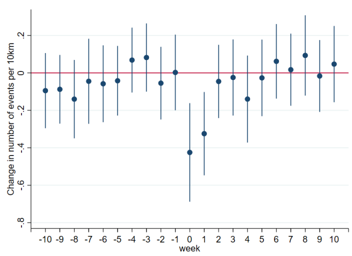
(a) Coach # 1



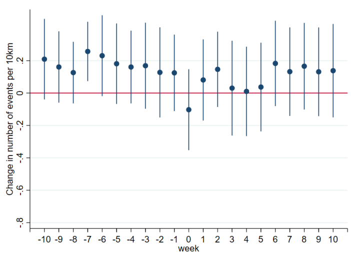
(b) Coach # 2



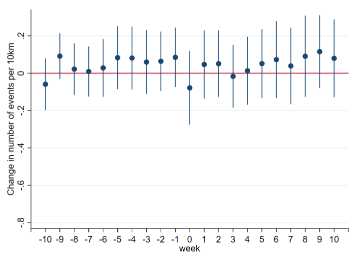
(c) Coach # 3



(d) Coach # 4



(e) Coach # 5



(f) Coach # 6

F Regression Estimates Temporal Effects of Coaching on Braking and Cornering

Table A14: Written Feedback on Driving Performance, Multiple Hypotheses Testing Correction

Outcome		Δ	*p*-values		
			Unadj.	Bonf.	Holm
Fuel Economy	Post-announcement	-0.3674	0.0000***	0.0000***	0.0000***
	Post-feedback	-0.1726	0.0053***	0.0635*	0.0423**
	Post-experiment	0.5529	0.0000***	0.0000***	0.0000***
Acceleration	Post-announcement	-1.2294	0.0000***	0.0000***	0.0000***
	Post-feedback	-0.6928	0.0000***	0.0000***	0.0000***
	Post-experiment	0.1866	0.0732	0.879	0.6592
Braking	Post-announcement	-1.2917	0.0000***	0.0000***	0.0000***
	Post-feedback	0.0005	0.9898	1.0000	1.0000
	Post-experiment	-0.0111	0.6855	1.0000	1.0000
Cornering	Post-announcement	-0.3168	0.0000***	0.0000***	0.0000***
	Post-feedback	-0.1931	0.0001***	0.0017***	0.001***
	Post-experiment	0.0629	0.3588	1.0000	1.0000

Notes: Identification of written feedback on driving performance. The dependent variable fuel economy is measured in liters/100km; acceleration, braking and cornering as the number of events per 10 kilometers. The time period under consideration is from 01/01/2015 until 31/01/2017. Standard *p*-values as well as *p*-values that use a Bonferroni and Holm correction for multiple hypothesis testing are reported. Standard errors are clustered by driver. Full regression results are reported in Tables 6 and Table A5.
***(**,*) : the corresponding *p*-values are less than 1% (5% or 10%).

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. The post-announcement dummy variable is one in the period from November 9, 2015, onwards (kickoff-event), zero otherwise. Drivers are considered to be in the post-feedback period when they have received at least one report in the past. For most drivers, this was at 15 December 2015 and after. The post-experimental period starts at November 15, 2016, when it was communicated to the treated drivers that they will no longer receive peer-comparison messages. The dependent variable fuel economy is measured in liters/100km and acceleration as the number of events per 10 kilometers. Standard errors are clustered by driver. Controls include: travel distance, route dummies, number of passengers and bus stops, and dummies for bus type, morning and evening rush hours, non-scheduled rides and having been coached. Weather dummies: daily temperature, wind and rainfall. A no-report indicator is included to capture drivers operating after December 15, 2015, but who have not yet received their first report.

Table A15: In-Person Coaching Effects on Driving Performance: Braking and Cornering

Dependent variable:	Braking			Cornering		
	(1)	(2)	(3)	(4)	(5)	(6)
Post-coaching	-0.009	-0.016		-0.047	-0.047	
	(0.026)	(0.025)		(0.038)	(0.038)	
Day of first coaching			-0.109***			-0.190***
			(0.034)			(0.039)
1 –7 days after			-0.024			-0.104***
			(0.025)			(0.035)
8 –14 days after			-0.063**			-0.027
			(0.027)			(0.035)
15 –21 days after			-0.016			-0.079**
			(0.025)			(0.037)
22 –28 days after			-0.031			-0.098**
			(0.026)			(0.040)
29 –35 days after			-0.018			-0.054
			(0.028)			(0.038)
36 –42 days after			-0.004			-0.020
			(0.029)			(0.045)
43 –49 days after			0.017			-0.039
			(0.031)			(0.047)
50 –56 days after			-0.007			0.012
			(0.031)			(0.051)
57 –63 days after			0.009			-0.030
			(0.030)			(0.051)
64 –70 days after			-0.005			-0.015
			(0.032)			(0.052)
> 70			0.020			0.018
			(0.037)			(0.055)
Number of trip-level observations	187127	187127	187127	187127	187127	187127
Controls	No	Yes	Yes	No	Yes	Yes
Driver fixed effects	No	Yes	Yes	No	Yes	Yes
Day fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Bus type × day fixed effects	No	Yes	Yes	No	Yes	Yes

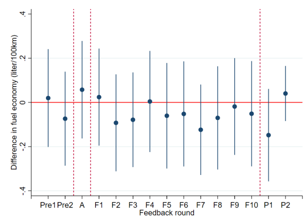
Notes: Identification of in-person coaching effects on driving performance. The time period under consideration is the period for which we have complete logs available from all coaches (01/01/2015-30/04/2016). Post-coaching identifies the day of first coaching and the full period thereafter for each coached driver. The dependent variables braking and cornering are measured as the number of events per 10 kilometers. Standard errors are clustered by driver. Controls include: travel distance, route dummies, number of passengers and bus stops, and dummies for bus type, morning and evening rush hours, fill-in rides and having been coached. Weather dummies: daily temperature, wind and rainfall. ***(**,*) : statistically different from zero at the 1%-level (5%-level, 10%-level). Standard errors in parentheses.

G Further Results: Targeted Peer-Comparison Feedback

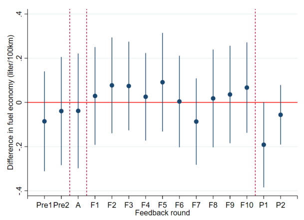
G.1 Intertemporal Treatment Differences

Figure A9 examines temporal effects by plotting the treatment effects per feedback round. The first round starts around 15 December 2015, with a new report being distributed in each subsequent month. The feedback report in November 2016 contains a text message notifying all treated drivers that they will no longer receive peer-comparison messages. The general pattern is that there are no intertemporal effects of the peer-comparison messages on driving behavior.

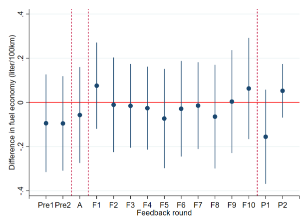
Figure A9: Intertemporal Treatment Differences Targeted peer-comparison Feedback [Fuel Economy and Acceleration]



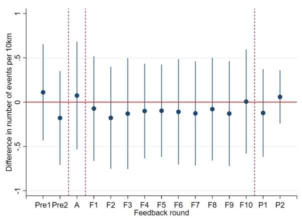
(a) T2-1n0p: fuel economy



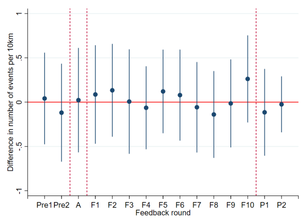
(b) T3-1n1p: fuel economy



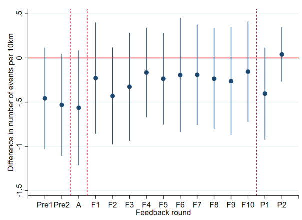
(c) T4-3n0p: fuel economy



(d) T2-1n0p: acceleration



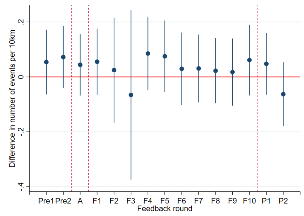
(e) T3-1n1p: acceleration



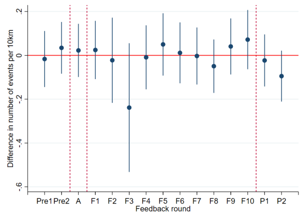
(f) T4-3n0p: acceleration

Notes: Treatment effects per feedback round based on trips with VDL and Intouro buses. The time period is from 01/09/2015-31/01/2017. **Pre:** 01/09/2015-08/11/2015 [pre-announcement period]; **A:** 09/11/2015-14/12/2015 [announcement period]; **F*i*:** 15/12/2015-14/11/2016 [feedback period]; **P*i*:** 15/11/2016-31/01/2017 [post-experiment period]. The dependent variable fuel economy is measured in liters/100km and acceleration as the number of events per 10 kilometers. Standard errors are clustered by driver. Controls include: travel distance, route dummies, number of passengers and bus stops, and dummies for bus type, morning and evening rush hours and fill-in rides. Weather dummies: daily temperature, wind and rainfall. A no-report indicator is included to capture drivers operating after 15/12/2015 (first feedback round) but who have not yet received their first report.

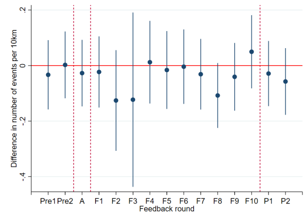
Figure A10: Intertemporal Treatment Differences Targeted Peer-Comparison Feedback [Braking and Cornering]



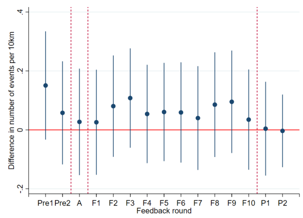
(a) T2-1n0p: braking



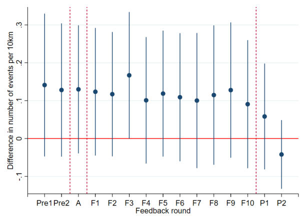
(b) T3-1n1p: braking



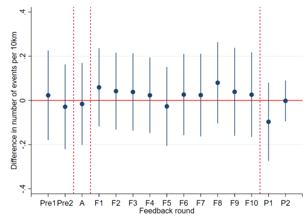
(c) T4-3n0p: braking



(d) T2-1n0p: cornering



(e) T3-1n1p: cornering



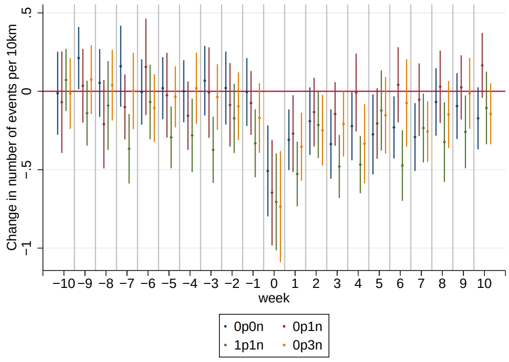
(f) T4-3n0p: cornering

Notes: Treatment effects per feedback round based on trips with VDL and Intouro buses. The time period is from 01/09/2015-31/01/2017. **Pre:** 01/09/2015-08/11/2015 [pre-announcement period]; **A:** 09/11/2015-14/12/2015 [announcement period]; **F*i*:** 15/12/2015-14/11/2016 [feedback period]; **P*i*:** 15/11/2016-31/01/2017 [post-experiment period]. The dependent variables braking and cornering are measured as the number of events per 10 kilometers. Standard errors are clustered by driver. Controls include: travel distance, route dummies, number of passengers and bus stops, and dummies for bus type, morning and evening rush hours and fill-in rides. Weather dummies: daily temperature, wind and rainfall. A no-report indicator is included to capture drivers operating after 15/12/2015 (first feedback round) but who have not yet received their first report.

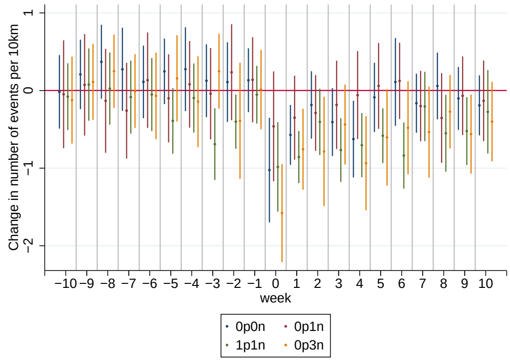
H Further Results on Treatment Complementarity

H.1 Effects In-Person Coaching, Conditional on Peer-Comparison Treatment

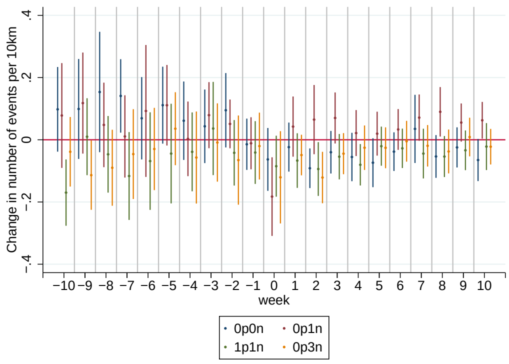
Figure A11: Treatment Level Effects In-Person Coaching: Fuel Economy and ABC



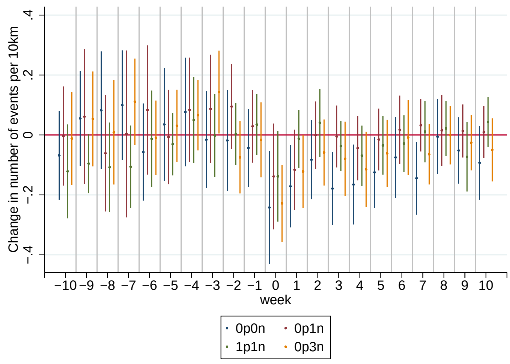
(a) Fuel economy



(b) Acceleration



(c) Braking



(d) Cornering

Notes: Driving performance in the 10 weeks before and after coaching based on trips with VDL and Intouro buses. The day of coaching itself is point 0 on the x-axis. The vertical spikes indicate 95% confidence intervals. The dependent variables braking and cornering are measured as the number of events per 10 kilometers. Standard errors are clustered by driver. Controls include: travel distance, route dummies, number of passengers and bus stops, and dummies for bus type, morning and evening rush hours and fill-in rides the interaction of bus type and date fixed effects. Coaches are excluded.

H.2 Effects Peer-Comparison Feedback, Conditional on Being Coached

Table A16: Targeted peer-comparison Feedback Effects on Driving Performance - Driver NOT COACHED Receiving First Feedback

Dependent variable:	**Fuel Economy**				**Acceleration**			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Post-announcement	-0.172	-0.409***	-0.338***		-1.693***	-0.974***	-1.090***	
	(0.121)	(0.097)	(0.068)		(0.231)	(0.200)	(0.175)	
T2 (1n/0p)	0.115	0.248			0.016	0.040		
	(0.232)	(0.220)			(0.420)	(0.426)		
T3 (1n/1p)	0.350	0.357			0.470	0.442		
	(0.349)	(0.339)			(0.636)	(0.596)		
T4 (3n/0p)	0.277	0.225			0.541	0.565		
	(0.329)	(0.297)			(0.524)	(0.530)		
Post-feedback	-0.371***	-0.036	-0.023		-0.611**	-0.539**	-0.335	
	(0.141)	(0.127)	(0.105)		(0.289)	(0.252)	(0.202)	
Post-feedback × T2 (1n/0p)	-0.426**	-0.412**	-0.379**	-0.337**	-0.269	-0.079	-0.409	-0.455
	(0.204)	(0.182)	(0.155)	(0.149)	(0.386)	(0.339)	(0.318)	(0.315)
Post-feedback × T3 (1n/1p)	0.001	-0.015	-0.117	-0.106	-0.370	-0.334	-0.512	-0.536
	(0.180)	(0.171)	(0.181)	(0.180)	(0.488)	(0.426)	(0.417)	(0.415)
Post-feedback × T4 (3n/0p)	-0.354**	-0.216	-0.304**	-0.284**	-0.889**	-0.945***	-1.012***	-1.024***
	(0.163)	(0.140)	(0.147)	(0.141)	(0.405)	(0.317)	(0.326)	(0.334)
Post-experiment	0.517***	0.519***	0.476***		0.371	0.166	0.161	
	(0.145)	(0.091)	(0.086)		(0.284)	(0.205)	(0.188)	
Post-experiment × T2 (1n/0p)	-0.048	0.046	0.122	0.110	-0.026	-0.009	0.024	0.019
	(0.207)	(0.127)	(0.130)	(0.115)	(0.341)	(0.274)	(0.235)	(0.243)
Post-experiment × T3 (1n/1p)	0.144	0.131	0.084	0.091	0.158	0.193	0.056	0.030
	(0.197)	(0.146)	(0.131)	(0.119)	(0.404)	(0.297)	(0.266)	(0.264)
Post-experiment × T4 (3n/0p)	-0.156	0.033	0.058	0.052	0.049	0.189	0.236	0.239
	(0.199)	(0.143)	(0.134)	(0.125)	(0.387)	(0.316)	(0.297)	(0.278)
Constant	24.196***	21.809***	22.397***	24.067***	10.283***	8.538***	11.257***	10.921***
	(0.181)	(0.207)	(0.161)	(0.386)	(0.357)	(0.361)	(0.376)	(0.418)
R^2	.0175	.409	.502	.526	.0451	.374	.519	.541
Number of trip-level observations	144305	136482	136482	136482	99562	94667	94667	94667
Controls	No	Yes	Yes	Yes	No	Yes	Yes	Yes
Weather dummies	No	Yes	Yes	No	No	Yes	Yes	No
Driver fixed effects	No	No	Yes	Yes	No	No	Yes	Yes
Day fixed effects	No	No	No	Yes	No	No	No	Yes
Bus type × day fixed effects	No	No	No	Yes	No	No	No	Yes

Notes: Identification of the treatment effects on driving performance. The time period under consideration is from 01/01/2015 until 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. The post-announcement dummy variable is one in the period from 09/11/2015 onwards (feedback announcement), zero otherwise. Drivers are considered to be in the post-feedback period when they have received at least one report in the past. For most drivers, this was at 15/12/2015 and after. The post-experimental period starts at 15/11/2016 when it was communicated to the treated drivers that they will no longer receive peer-comparison messages. The dependent variable fuel economy is measured in liters/100km and acceleration as the number of events per 10 kilometers. Standard errors are clustered by driver. Controls include: travel distance, route dummies, number of passengers and bus stops, and dummies for bus type, morning and evening rush hours and fill-in rides. Weather dummies: daily temperature, wind and rainfall. A no-report indicator is included to capture drivers operating after 15/12/2015 (first feedback round) but who have not yet received their first report.
***(**,*) : statistically different from zero at the 1%-level (5%-level, 10%-level). Standard errors in parentheses.

Table A17: Targeted Peer-Comparison Feedback Effects on Driving Performance - Drivers COACHED Before First Feedback

Dependent variable:	Fuel Economy				Acceleration			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Post-announcement	-0.452***	-0.392***	-0.455***		-2.617***	-1.069***	-1.261***	
	(0.091)	(0.105)	(0.072)		(0.180)	(0.194)	(0.168)	
T2 (1n/0p)	-0.552**	-0.518***			-0.510	-0.298		
	(0.266)	(0.194)			(0.497)	(0.404)		
T3 (1n/1p)	-0.332	-0.313			0.023	-0.046		
	(0.255)	(0.207)			(0.582)	(0.463)		
T4 (3n/0p)	-0.526**	-0.478**			-0.332	-0.155		
	(0.254)	(0.184)			(0.479)	(0.397)		
Post-feedback	-0.477***	0.007	-0.171*		-0.436*	-0.400	-0.673***	
	(0.123)	(0.158)	(0.100)		(0.242)	(0.291)	(0.249)	
Post-feedback × T2 (1n/0p)	0.090	0.071	0.117	0.120	0.404	0.349	0.260	0.238
	(0.212)	(0.154)	(0.135)	(0.135)	(0.411)	(0.354)	(0.356)	(0.351)
Post-feedback × T3 (1n/1p)	0.165	0.134	0.082	0.069	0.004	0.185	0.183	0.200
	(0.152)	(0.110)	(0.107)	(0.106)	(0.342)	(0.288)	(0.284)	(0.282)
Post-feedback × T4 (3n/0p)	0.173	0.126	0.064	0.035	0.059	0.286	0.242	0.254
	(0.174)	(0.109)	(0.101)	(0.097)	(0.321)	(0.262)	(0.246)	(0.246)
Post-experiment	0.547***	0.486***	0.537***		0.308	0.117	0.220	
	(0.144)	(0.093)	(0.082)		(0.235)	(0.196)	(0.163)	
Post-experiment × T2 (1n/0p)	0.235	0.238	0.179	0.186	0.516	0.465	0.316	0.307
	(0.216)	(0.169)	(0.149)	(0.140)	(0.352)	(0.321)	(0.273)	(0.266)
Post-experiment × T3 (1n/1p)	0.134	0.166	0.089	0.107	0.431	0.427*	0.256	0.274
	(0.185)	(0.127)	(0.105)	(0.107)	(0.296)	(0.249)	(0.216)	(0.216)
Post-experiment × T4 (3n/0p)	0.014	0.156	0.056	0.056	0.156	0.253	0.075	0.087
	(0.171)	(0.114)	(0.104)	(0.105)	(0.313)	(0.251)	(0.193)	(0.192)
Constant	24.949***	22.243***	22.581***	23.768***	10.914***	8.389***	11.638***	11.793***
	(0.181)	(0.142)	(0.125)	(0.289)	(0.387)	(0.315)	(0.268)	(0.377)
R^2	.0183	.424	.498	.517	.0458	.414	.561	.578
Number of trip-level observations	243265	232597	232597	232597	171092	164593	164593	164593
Controls	No	Yes	Yes	Yes	No	Yes	Yes	Yes
Weather dummies	No	Yes	Yes	No	No	Yes	Yes	No
Driver fixed effects	No	No	Yes	Yes	No	No	Yes	Yes
Day fixed effects	No	No	No	Yes	No	No	No	Yes
Bus type × day fixed effects	No	No	No	Yes	No	No	No	Yes

Notes: Identification of the treatment effects on driving performance. The time period under consideration is from 01/01/2015 until 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. The post-announcement dummy variable is one in the period from 09/11/2015 onwards (feedback announcement), zero otherwise. Drivers are considered to be in the post-feedback period when they have received at least one report in the past. For most drivers, this was at 15/12/2015 and after. The post-experimental period starts at 15/11/2016 when it was communicated to the treated drivers that they will no longer receive peer-comparison messages. The dependent variable fuel economy is measured in liters/100km and acceleration as the number of events per 10 kilometers. Standard errors are clustered by driver. Controls include: travel distance, route dummies, number of passengers and bus stops, and dummies for bus type, morning and evening rush hours, fill-in rides and having been coached. Weather dummies: daily temperature, wind and rainfall. A no-report indicator is included to capture drivers operating after 15/12/2015 (first feedback round) but who have not yet received their first report.
***(**,*) : statistically different from zero at the 1%-level (5%-level, 10%-level). Standard errors in parentheses.

Table A18: Targeted Peer-Comparison Feedback Effects on Driving Performance - Drivers NOT COACHED Before First Feedback

Dependent variable:	**Braking**				**Cornering**			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Post-announcement	-1.264***	-1.294***	-1.302***		-0.249***	-0.255***	-0.328***	
	(0.047)	(0.052)	(0.050)		(0.072)	(0.067)	(0.059)	
T2 (1n/0p)	0.046	0.065			0.197	0.195		
	(0.106)	(0.104)			(0.235)	(0.235)		
T3 (1n/1p)	-0.016	-0.004			-0.022	-0.038		
	(0.125)	(0.122)			(0.245)	(0.234)		
T4 (3n/0p)	0.203	0.222*			0.392	0.391		
	(0.124)	(0.122)			(0.262)	(0.258)		
Post-feedback	0.182***	0.199***	0.199***		-0.115	-0.094	-0.098	
	(0.051)	(0.053)	(0.057)		(0.120)	(0.115)	(0.107)	
Post-feedback × T2 (1n/0p)	-0.081	-0.116*	-0.130*	-0.144*	-0.085	-0.070	-0.047	-0.061
	(0.063)	(0.068)	(0.073)	(0.078)	(0.120)	(0.118)	(0.116)	(0.118)
Post-feedback × T3 (1n/1p)	-0.085	-0.097	-0.088	-0.091	0.067	0.077	0.081	0.074
	(0.102)	(0.099)	(0.106)	(0.103)	(0.131)	(0.121)	(0.119)	(0.120)
Post-feedback × T4 (3n/0p)	-0.280***	-0.291***	-0.282***	-0.279***	-0.318**	-0.324**	-0.226	-0.230
	(0.090)	(0.091)	(0.090)	(0.092)	(0.150)	(0.148)	(0.137)	(0.139)
Post-experiment	0.006	-0.021	-0.003		-0.217***	-0.221***	-0.092*	
	(0.056)	(0.048)	(0.042)		(0.072)	(0.074)	(0.051)	
Post-experiment × T2 (1n/0p)	0.066	0.034	0.026	0.040	0.191**	0.172*	0.059	0.046
	(0.063)	(0.061)	(0.053)	(0.055)	(0.092)	(0.089)	(0.064)	(0.061)
Post-experiment × T3 (1n/1p)	0.104	0.132*	0.099	0.113	0.117	0.113	0.002	-0.009
	(0.084)	(0.073)	(0.071)	(0.069)	(0.089)	(0.091)	(0.069)	(0.068)
Post-experiment × T4 (3n/0p)	0.086	0.078	0.057	0.060	0.193*	0.181	0.078	0.079
	(0.068)	(0.061)	(0.060)	(0.057)	(0.111)	(0.112)	(0.095)	(0.093)
Constant	1.641***	1.362***	2.261***	1.991***	1.039***	1.162***	1.668***	1.615***
	(0.091)	(0.103)	(0.112)	(0.174)	(0.188)	(0.202)	(0.107)	(0.133)
R^2	.0833	.196	.219	.29	.0306	.0973	.407	.421
Number of trip-level observations	99562	94667	94667	94667	100837	95938	95938	95938
Controls	No	Yes	Yes	Yes	No	Yes	Yes	Yes
Weather dummies	No	Yes	Yes	No	No	Yes	Yes	No
Driver fixed effects	No	No	Yes	Yes	No	No	Yes	Yes
Day fixed effects	No	No	No	Yes	No	No	No	Yes
Bus type × day fixed effects	No	No	No	Yes	No	No	No	Yes

Notes: Identification of the treatment effects on driving performance. The time period under consideration is from 01/01/2015 until 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. The post-announcement dummy variable is one in the period from 09/11/2015 onwards (feedback announcement), zero otherwise. Drivers are considered to be in the post-feedback period when they have received at least one report in the past. For most drivers, this was at 15/12/2015 and after. The post-experimental period starts at 15/11/2016 when it was communicated to the treated drivers that they will no longer receive peer-comparison messages. The dependent variable fuel economy is measured in liters/100km and acceleration as the number of events per 10 kilometers. Standard errors are clustered by driver. Controls include: travel distance, route dummies, number of passengers and bus stops, and dummies for bus type, morning and evening rush hours and fill-in rides. Weather dummies: daily temperature, wind and rainfall. A no-report indicator is included to capture drivers operating after 15/12/2015 (first feedback round) but who have not yet received their first report.
***(**,*) : statistically different from zero at the 1%-level (5%-level, 10%-level). Standard errors in parentheses.

Table A19: Targeted Peer-Comparison Feedback Effects on Driving Performance - Drivers COACHED Before First Feedback

Dependent variable:	**Braking**				**Cornering**			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Post-announcement	-1.248***	-1.321***	-1.336***		-0.316***	-0.120*	-0.225***	
	(0.043)	(0.049)	(0.048)		(0.038)	(0.069)	(0.051)	
T2 (1n/0p)	-0.124	-0.111			-0.196	-0.172		
	(0.091)	(0.086)			(0.191)	(0.176)		
T3 (1n/1p)	-0.145	-0.138			-0.228	-0.235		
	(0.092)	(0.089)			(0.175)	(0.159)		
T4 (3n/0p)	-0.167	-0.168*			-0.167	-0.125		
	(0.106)	(0.097)			(0.169)	(0.166)		
Post-feedback	0.017	-0.156***	-0.159***		-0.122**	-0.056	-0.157**	
	(0.048)	(0.051)	(0.051)		(0.052)	(0.086)	(0.066)	
Post-feedback × T2 (1n/0p)	0.135*	0.112	0.101	0.087	0.029	0.026	0.061	0.054
	(0.076)	(0.071)	(0.071)	(0.067)	(0.115)	(0.109)	(0.111)	(0.110)
Post-feedback × T3 (1n/1p)	0.069	0.055	0.056	0.055	0.052	0.065	0.075	0.070
	(0.073)	(0.068)	(0.068)	(0.066)	(0.090)	(0.085)	(0.090)	(0.088)
Post-feedback × T4 (3n/0p)	0.113	0.122	0.095	0.097	0.149**	0.148**	0.139*	0.130*
	(0.082)	(0.079)	(0.076)	(0.074)	(0.074)	(0.073)	(0.075)	(0.074)
Post-experiment	0.134**	0.060	0.066		-0.066	-0.092	-0.026	
	(0.053)	(0.044)	(0.045)		(0.064)	(0.056)	(0.035)	
Post-experiment × T2 (1n/0p)	0.007	-0.011	-0.008	-0.017	0.047	0.067	-0.021	-0.024
	(0.071)	(0.063)	(0.062)	(0.061)	(0.081)	(0.071)	(0.054)	(0.054)
Post-experiment × T3 (1n/1p)	0.036	0.055	0.039	0.029	0.014	0.032	-0.033	-0.035
	(0.071)	(0.053)	(0.053)	(0.054)	(0.071)	(0.058)	(0.042)	(0.042)
Post-experiment × T4 (3n/0p)	0.030	0.068	0.055	0.042	0.052	0.067	-0.039	-0.040
	(0.065)	(0.064)	(0.062)	(0.059)	(0.081)	(0.069)	(0.051)	(0.050)
Constant	1.786***	1.392***	2.264***	2.113***	1.178***	1.304***	1.492***	1.643***
	(0.079)	(0.083)	(0.093)	(0.119)	(0.131)	(0.134)	(0.075)	(0.117)
R^2	.0767	.206	.233	.291	.0235	.101	.379	.391
Number of trip-level observations	171092	164593	164593	164593	174143	167632	167632	167632
Controls	No	Yes	Yes	Yes	No	Yes	Yes	Yes
Weather dummies	No	Yes	Yes	No	No	Yes	Yes	No
Driver fixed effects	No	No	Yes	Yes	No	No	Yes	Yes
Day fixed effects	No	No	No	Yes	No	No	No	Yes
Bus type × day fixed effects	No	No	No	Yes	No	No	No	Yes

Notes: Identification of the treatment effects on driving performance. The time period under consideration is from 01/01/2015 until 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. The post-announcement dummy variable is one in the period from 09/11/2015 onwards (feedback announcement), zero otherwise. Drivers are considered to be in the post-feedback period when they have received at least one report in the past. For most drivers, this was at 15/12/2015 and after. The post-experimental period starts at 15/11/2016 when it was communicated to the treated drivers that they will no longer receive peer-comparison messages. The dependent variable fuel economy is measured in liters/100km and acceleration as the number of events per 10 kilometers. Standard errors are clustered by driver. Controls include: travel distance, route dummies, number of passengers and bus stops, and dummies for bus type, morning and evening rush hours, fill-in rides and having been coached. Weather dummies: daily temperature, wind and rainfall. A no-report indicator is included to capture drivers operating after 15/12/2015 (first feedback round) but who have not yet received their first report.
***(**,*) : statistically different from zero at the 1%-level (5%-level, 10%-level). Standard errors in parentheses.

Table A20: Targeted Peer-Comparison Feedback Effects on Driving Performance [Groups: Age]

Dependent variable:	**Fuel economy**	**Acceleration**	**Braking**	**Cornering**
	(1)	(2)	(3)	(4)
Post-announcement=1	-0.371***	-0.911***	-1.214***	-0.287***
	(0.065)	(0.185)	(0.049)	(0.054)
Post-announcement=1 × agegroup=1	-0.310***	-0.800***	-0.093	-0.082
	(0.114)	(0.280)	(0.069)	(0.087)
Post-announcement=1 × agegroup=2	0.016	0.191	-0.135*	-0.063
	(0.105)	(0.251)	(0.081)	(0.075)
Post-announcement=1 × agegroup=4	0.047	-0.011	-0.023	-0.074
	(0.092)	(0.230)	(0.071)	(0.075)
Post-feedback=1	-0.132**	-0.972***	0.008	-0.210***
	(0.065)	(0.172)	(0.028)	(0.044)
Post-feedback=1 × agegroup=1	-0.095	0.113	-0.026	0.062
	(0.112)	(0.270)	(0.040)	(0.057)
Post-feedback=1 × agegroup=2	-0.073	0.049	-0.027	0.043
	(0.125)	(0.300)	(0.046)	(0.083)
Post-feedback=1 × agegroup=4	0.069	0.391*	0.030	0.167***
	(0.103)	(0.213)	(0.044)	(0.054)
Post-experiment=1	0.599***	0.436***	0.070***	-0.024
	(0.043)	(0.106)	(0.024)	(0.026)
Post-experiment=1 × agegroup=1	-0.046	-0.244	-0.077**	-0.026
	(0.076)	(0.164)	(0.035)	(0.035)
Post-experiment=1 × agegroup=2	-0.033	-0.137	-0.010	-0.071
	(0.078)	(0.183)	(0.037)	(0.053)
Post-experiment=1 × agegroup=4	0.014	-0.103	0.003	-0.067*
	(0.067)	(0.174)	(0.034)	(0.035)
Constant	22.555***	11.357***	2.314***	1.642***
	(0.078)	(0.188)	(0.059)	(0.052)
R^2	.531	.45	.198	.396
# trip-level observations	484918	349879	349879	349879
Controls	Yes	Yes	Yes	Yes
Weather dummies	Yes	Yes	Yes	Yes
Driver fixed effects	Yes	Yes	Yes	Yes
Day fixed effects	No	No	No	No
Bus type × day fixed effects	No	No	No	No

Notes: Identification of the treatment effects on driving performance. The time period under consideration is from 01/01/2015 until 31/01/2017. The post-announcement dummy variable is one in the period from 09/11/2015 onwards (feedback announcement), zero otherwise. Drivers are considered to be in the post-feedback period when they have received at least one report in the past. For most drivers, this was at 15/12/2015 and after. The post-experimental period starts at 15/11/2016 when it was communicated to the treated drivers that they will no longer receive peer-comparison messages. The dependent variable fuel economy is measured in liters/100km and acceleration as the number of events per 10 kilometers. Standard errors are clustered by driver. Controls include: travel distance, route dummies, number of passengers and bus stops, and dummies for bus type, morning and evening rush hours, fill-in rides and having been coached. Weather dummies: daily temperature, wind and rainfall. A no-report indicator is included to capture drivers operating after 15/12/2015 (first feedback round) but who have not yet received their first report. Default agegroup=3 (55-59 years); agegroup=1: < 50 years; agegroup=2: $50 - 54$ years; agegroup=4: ≥ 60 years.
***(**,*) : statistically different from zero at the 1%-level (5%-level, 10%-level). Standard errors in parentheses.

Table A21: Targeted Peer-Comparison Feedback Effects on Driving Performance [Groups: Years of Service]

Dependent variable:	**Fuel economy**	**Acceleration**	**Braking**	**Cornering**
	(1)	(2)	(3)	(4)
Post-announcement=1	-0.490***	-1.193***	-1.254***	-0.342***
	(0.075)	(0.182)	(0.042)	(0.056)
Post-announcement=1 × dienstgroup=1	0.017	0.082	-0.016	0.014
	(0.101)	(0.245)	(0.070)	(0.076)
Post-announcement=1 × dienstgroup=2	0.016	0.073	-0.099	-0.008
	(0.120)	(0.280)	(0.070)	(0.092)
Post-announcement=1 × dienstgroup=4	0.180*	0.304	0.051	0.016
	(0.106)	(0.257)	(0.068)	(0.076)
Post-feedback=1	-0.155*	-0.938***	-0.008	-0.130***
	(0.090)	(0.193)	(0.029)	(0.043)
Post-feedback=1 × dienstgroup=1	-0.082	0.041	0.056	-0.006
	(0.117)	(0.273)	(0.042)	(0.064)
Post-feedback=1 × dienstgroup=2	0.006	0.227	-0.026	-0.066
	(0.129)	(0.247)	(0.046)	(0.062)
Post-feedback=1 × dienstgroup=4	0.129	0.242	0.016	0.012
	(0.110)	(0.248)	(0.045)	(0.058)
Post-experiment=1	0.637***	0.412***	0.049*	-0.026
	(0.054)	(0.128)	(0.025)	(0.033)
Post-experiment=1 × dienstgroup=1	-0.055	-0.068	-0.008	-0.042
	(0.078)	(0.178)	(0.035)	(0.042)
Post-experiment=1 × dienstgroup=2	-0.073	-0.158	-0.022	-0.059
	(0.084)	(0.190)	(0.039)	(0.044)
Post-experiment=1 × dienstgroup=4	-0.084	-0.178	0.019	-0.043
	(0.070)	(0.179)	(0.038)	(0.043)
Constant	22.562***	11.359***	2.318***	1.636***
	(0.080)	(0.190)	(0.058)	(0.053)
R^2	.531	.449	.198	.396
# trip-level observations	484918	349879	349879	349879
Controls	Yes	Yes	Yes	Yes
Weather dummies	Yes	Yes	Yes	Yes
Driver fixed effects	Yes	Yes	Yes	Yes
Day fixed effects	No	No	No	No
Bus type × day fixed effects	No	No	No	No

Notes: Identification of the treatment effects on driving performance. The time period under consideration is from 01/01/2015 until 31/01/2017. The post-announcement dummy variable is one in the period from 09/11/2015 onwards (feedback announcement), zero otherwise. Drivers are considered to be in the post-feedback period when they have received at least one report in the past. For most drivers, this was at 15/12/2015 and after. The post-experimental period starts at 15/11/2016 when it was communicated to the treated drivers that they will no longer receive peer-comparison messages. The dependent variable fuel economy is measured in liters/100km and acceleration as the number of events per 10 kilometers. Standard errors are clustered by driver. Controls include: travel distance, route dummies, number of passengers and bus stops, and dummies for bus type, morning and evening rush hours, fill-in rides and having been coached. Weather dummies: daily temperature, wind and rainfall. A no-report indicator is included to capture drivers operating after 15/12/2015 (first feedback round) but who have not yet received their first report. Default tenuregroup=3 (16-29 years); tenuregroup=1: < 8 years; tenuregroup=2: $8 - 15$ years; tenuregroup=4: ≥ 30 years.
***(**,*) : statistically different from zero at the 1%-level (5%-level, 10%-level). Standard errors in parentheses.

Table A22: Targeted Peer-Comparison Feedback Effects on Driving Performance [Groups: Gender]

Dependent variable:	**Fuel economy**	**Acceleration**	**Braking**	**Cornering**
	(1)	(2)	(3)	(4)
Post-announcement=1	-0.429***	-1.100***	-1.269***	-0.341***
	(0.044)	(0.104)	(0.028)	(0.032)
Post-announcement=1 × gendergroup=1	-0.063	0.155	0.032	0.038
	(0.090)	(0.266)	(0.098)	(0.086)
Post-feedback=1	-0.123**	-0.750***	0.008	-0.134***
	(0.049)	(0.103)	(0.018)	(0.028)
Post-feedback=1 × gendergroup=1	-0.199	-0.625	-0.044	-0.058
	(0.154)	(0.397)	(0.045)	(0.068)
Post-experiment=1	0.612***	0.369***	0.055***	-0.059***
	(0.028)	(0.066)	(0.013)	(0.015)
Post-experiment=1 × gendergroup=1	-0.298**	-0.580**	-0.094*	-0.032
	(0.119)	(0.229)	(0.049)	(0.044)
Constant	22.554***	11.368***	2.317***	1.637***
	(0.079)	(0.193)	(0.059)	(0.054)
R^2	.529	.448	.197	.393
Number of trip-level observations	484918	349879	349879	349879
Controls	Yes	Yes	Yes	Yes
Weather dummies	Yes	Yes	Yes	Yes
Driver fixed effects	Yes	Yes	Yes	Yes
Day fixed effects	No	No	No	No
Bus type × day fixed effects	No	No	No	No

Notes: Identification of the treatment effects on driving performance. The time period under consideration is from 01/01/2015 until 31/01/2017. The post-announcement dummy variable is one in the period from 09/11/2015 onwards (feedback announcement), zero otherwise. Drivers are considered to be in the post-feedback period when they have received at least one report in the past. For most drivers, this was at 15/12/2015 and after. The post-experimental period starts at 15/11/2016 when it was communicated to the treated drivers that they will no longer receive peer-comparison messages. The dependent variable fuel economy is measured in liters/100km and acceleration as the number of events per 10 kilometers. Standard errors are clustered by driver. Controls include: travel distance, route dummies, number of passengers and bus stops, and dummies for bus type, morning and evening rush hours, fill-in rides and having been coached. Weather dummies: daily temperature, wind and rainfall. A no-report indicator is included to capture drivers operating after 15/12/2015 (first feedback round) but who have not yet received their first report. Default gendergroup=0 (males).

***(**,*) : statistically different from zero at the 1%-level (5%-level, 10%-level). Standard errors in parentheses.

List of research reports

15001-EEF: Bao, T., X. Tian, X. Yu, Dictator Game with Indivisibility of Money

15002-GEM: Chen, Q., E. Dietzenbacher, and B. Los, The Effects of Ageing and Urbanization on China's Future Population and Labor Force

15003-EEF: Allers, M., B. van Ommeren, and B. Geertsema, Does intermunicipal cooperation create inefficiency? A comparison of interest rates paid by intermunicipal organizations, amalgamated municipalities and not recently amalgamated municipalities

15004-EEF: Dijkstra, P.T., M.A. Haan, and M. Mulder, Design of Yardstick Competition and Consumer Prices: Experimental Evidence

15005-EEF: Dijkstra, P.T., Price Leadership and Unequal Market Sharing: Collusion in Experimental Markets

15006-EEF: Anufriev, M., T. Bao, A. Sutin, and J. Tuinstra, Fee Structure, Return Chasing and Mutual Fund Choice: An Experiment

15007-EEF: Lamers, M., Depositor Discipline and Bank Failures in Local Markets During the Financial Crisis

15008-EEF: Oosterhaven, J., On de Doubtful Usability of the Inoperability IO Model

15009-GEM: Zhang, L. and D. Bezemer, A Global House of Debt Effect? Mortgages and Post-Crisis Recessions in Fifty Economies

15010-I&O: Hooghiemstra, R., N. Hermes, L. Oxelheim, and T. Randøy, The Impact of Board Internationalization on Earnings Management

15011-EEF: Haan, M.A., and W.H. Siekman, Winning Back the Unfaithful while Exploiting the Loyal: Retention Offers and Heterogeneous Switching Costs

15012-EEF: Haan, M.A., J.L. Moraga-González, and V. Petrikaite, Price and Match-Value Advertising with Directed Consumer Search

15013-EEF: Wiese, R., and S. Eriksen, Do Healthcare Financing Privatisations Curb Total Healthcare Expenditures? Evidence from OECD Countries

15014-EEF: Siekman, W.H., Directed Consumer Search

15015-GEM: Hoorn, A.A.J. van, Organizational Culture in the Financial Sector: Evidence from a Cross-Industry Analysis of Employee Personal Values and Career Success

15016-EEF: Te Bao, and C. Hommes, When Speculators Meet Constructors: Positive and Negative Feedback in Experimental Housing Markets

15017-EEF: Te Bao, and Xiaohua Yu, Memory and Discounting: Theory and Evidence

15018-EEF: Suari-Andreu, E., The Effect of House Price Changes on Household Saving Behaviour: A Theoretical and Empirical Study of the Dutch Case

15019-EEF: Bijlsma, M., J. Boone, and G. Zwart, Community Rating in Health Insurance: Trade-off between Coverage and Selection

15020-EEF: Mulder, M., and B. Scholtens, A Plant-level Analysis of the Spill-over Effects of the German *Energiewende*

15021-GEM: Samarina, A., L. Zhang, and D. Bezemer, Mortgages and Credit Cycle Divergence in Eurozone Economies

16001-GEM: Hoorn, A. van, How Are Migrant Employees Manages? An Integrated Analysis

16002-EEF: Soetevent, A.R., Te Bao, A.L. Schippers, A Commercial Gift for Charity

16003-GEM: Bouwmeerster, M.C., and J. Oosterhaven, Economic Impacts of Natural Gas Flow Disruptions

16004-MARK: Holtrop, N., J.E. Wieringa, M.J. Gijsenberg, and P. Stern, Competitive Reactions to Personal Selling: The Difference between Strategic and Tactical Actions

16005-EEF: Plantinga, A. and B. Scholtens, The Financial Impact of Divestment from Fossil Fuels

16006-GEM: Hoorn, A. van, Trust and Signals in Workplace Organization: Evidence from Job Autonomy Differentials between Immigrant Groups

16007-EEF: Willems, B. and G. Zwart, Regulatory Holidays and Optimal Network Expansion

16008-GEF: Hoorn, A. van, Reliability and Validity of the Happiness Approach to Measuring Preferences

16009-EEF: Hinloopen, J., and A.R. Soetevent, (Non-)Insurance Markets, Loss Size Manipulation and Competition: Experimental Evidence

16010-EEF: Bekker, P.A., A Generalized Dynamic Arbitrage Free Yield Model

16011-EEF: Mierau, J.A., and M. Mink, A Descriptive Model of Banking and Aggregate Demand

16012-EEF: Mulder, M. and B. Willems, Competition in Retail Electricity Markets: An Assessment of Ten Year Dutch Experience

16013-GEM: Rozite, K., D.J. Bezemer, and J.P.A.M. Jacobs, Towards a Financial Cycle for the US, 1873-2014

16014-EEF: Neuteleers, S., M. Mulder, and F. Hindriks, Assessing Fairness of Dynamic Grid Tariffs

16015-EEF: Soetevent, A.R., and T. Bružikas, Risk and Loss Aversion, Price Uncertainty and the Implications for Consumer Search

16016-HRM&OB: Meer, P.H. van der, and R. Wielers, Happiness, Unemployment and Self-esteem

16017-EEF: Mulder, M., and M. Pangan, Influence of Environmental Policy and Market Forces on Coal-fired Power Plants: Evidence on the Dutch Market over 2006-2014

16018-EEF: Zeng,Y., and M. Mulder, Exploring Interaction Effects of Climate Policies: A Model Analysis of the Power Market

16019-EEF: Ma, Yiqun, Demand Response Potential of Electricity End-users Facing Real Time Pricing

16020-GEM: Bezemer, D., and A. Samarina, Debt Shift, Financial Development and Income Inequality in Europe

16021-EEF: Elkhuizen, L, N. Hermes, and J. Jacobs, Financial Development, Financial Liberalization and Social Capital

16022-GEM: Gerritse, M., Does Trade Cause Institutional Change? Evidence from Countries South of the Suez Canal

16023-EEF: Rook, M., and M. Mulder, Implicit Premiums in Renewable-Energy Support Schemes

17001-EEF: Trinks, A., B. Scholtens, M. Mulder, and L. Dam, Divesting Fossil Fuels: The Implications for Investment Portfolios

17002-EEF: Angelini, V., and J.O. Mierau, Late-life Health Effects of Teenage Motherhood

17003-EEF: Jong-A-Pin, R., M. Laméris, and H. Garretsen, Political Preferences of (Un)happy Voters: Evidence Based on New Ideological Measures

17004-EEF: Jiang, X., N. Hermes, and A. Meesters, Financial Liberalization, the Institutional Environment and Bank Efficiency

17005-EEF: Kwaak, C. van der, Financial Fragility and Unconventional Central Bank Lending Operations

17006-EEF: Postelnicu, L. and N. Hermes, The Economic Value of Social Capital

17007-EEF: Ommeren, B.J.F. van, M.A. Allers, and M.H. Vellekoop, Choosing the Optimal Moment to Arrange a Loan

17008-EEF: Bekker, P.A., and K.E. Bouwman, A Unified Approach to Dynamic Mean-Variance Analysis in Discrete and Continuous Time

17009-EEF: Bekker, P.A., Interpretable Parsimonious Arbitrage-free Modeling of the Yield Curve

17010-GEM: Schasfoort, J., A. Godin, D. Bezemer, A. Caiani, and S. Kinsella, Monetary Policy Transmission in a Macroeconomic Agent-Based Model

17011-I&O: Bogt, H. ter, Accountability, Transparency and Control of Outsourced Public Sector Activities

17012-GEM: Bezemer, D., A. Samarina, and L. Zhang, The Shift in Bank Credit Allocation: New Data and New Findings

17013-EEF: Boer, W.I.J. de, R.H. Koning, and J.O. Mierau, Ex-ante and Ex-post Willingness-to-pay for Hosting a Major Cycling Event

17014-OPERA: Laan, N. van der, W. Romeijnders, and M.H. van der Vlerk, Higher-order Total Variation Bounds for Expectations of Periodic Functions and Simple Integer Recourse Approximations

17015-GEM: Oosterhaven, J., Key Sector Analysis: A Note on the Other Side of the Coin

17016-EEF: Romensen, G.J., A.R. Soetevent: Tailored Feedback and Worker Green Behavior: Field Evidence from Bus Drivers

17017-EEF: Trinks, A., G. Ibikunle, M. Mulder, and B. Scholtens, Greenhouse Gas Emissions Intensity and the Cost of Capital

17018-GEM: Qian, X. and A. Steiner, The Reinforcement Effect of International Reserves for Financial Stability

17019-GEM/EEF: Klasing, M.J. and P. Milionis, The International Epidemiological Transition and the Education Gender Gap

2018001-EEF: Keller, J.T., G.H. Kuper, and M. Mulder, Mergers of Gas Markets Areas and Competition amongst Transmission System Operators: Evidence on Booking Behaviour in the German Markets

2018002-EEF: Soetevent, A.R. and S. Adikyan, The Impact of Short-Term Goals on Long-Term Objectives: Evidence from Running Data

2018003-MARK: Gijsenberg, M.J. and P.C. Verhoef, Moving Forward: The Role of Marketing in Fostering Public Transport Usage

2018004-MARK: Gijsenberg, M.J. and V.R. Nijs, Advertising Timing: In-Phase or Out-of-Phase with Competitors?

2018005-EEF: Hulshof, D., C. Jepma, and M. Mulder, Performance of Markets for European Renewable Energy Certificates

2018006-EEF: Fosgaard, T.R., and A.R. Soetevent, Promises Undone: How Committed Pledges Impact Donations to Charity

2018007-EEF: Durán, N. and J.P. Elhorst, A Spatio-temporal-similarity and Common Factor Approach of Individual Housing Prices: The Impact of Many Small Earthquakes in the North of Netherlands

2018008-EEF: Hermes, N., and M. Hudon, Determinants of the Performance of Microfinance Institutions: A Systematic Review

2018009-EEF: Katz, M., and C. van der Kwaak, The Macroeconomic Effectiveness of Bank Bail-ins

2018010-OPERA: Prak, D., R.H. Teunter, M.Z. Babai, A.A. Syntetos, and J.E. Boylan, Forecasting and Inventory Control with Compound Poisson Demand Using Periodic Demand Data

2018011-EEF: Brock, B. de, Converting a Non-trivial Use Case into an SSD: An Exercise

2018012-EEF: Harvey, L.A., J.O. Mierau, and J. Rockey, Inequality in an Equal Society

2018013-OPERA: Romeijnders, W., and N. van der Laan, Inexact cutting planes for two-stage mixed-integer stochastic programs

2018014-EEF: Green, C.P., and S. Homroy, Bringing Connections Onboard: The Value of Political Influence

2018015-OPERA: Laan, N. van der, and W. Romeijnders, Generalized aplha-approximations for two-stage mixed-integer recourse models

2018016-GEM: Rozite, K., Financial and Real Integration between Mexico and the United States

2019001-EEF: Lugalla, I.M., J. Jacobs, and W. Westerman, Drivers of Women Entrepreneurs in Tourism in Tanzania: Capital, Goal Setting and Business Growth

2019002-EEF: Brock, E.O. de, On Incremental and Agile Development of (Information) Systems

2019003-OPERA: Laan, N. van der, R.H. Teunter, W. Romeijnders, and O.A. Kilic, The Data-driven Newsvendor Problem: Achieving On-target Service Levels.

2019004-EEF: Dijk, H., and J. Mierau, Mental Health over the Life Course: Evidence for a U-Shape?

2019005-EEF: Freriks, R.D., and J.O. Mierau, Heterogeneous Effects of School Resources on Child Mental Health Development: Evidence from the Netherlands.

2019006-OPERA: Broek, M.A.J. uit het, R.H. Teunter, B. de Jonge, J. Veldman, Joint Condition-based Maintenance and Condition-based Production Optimization.

2019007-OPERA: Broek, M.A.J. uit het, R.H. Teunter, B. de Jonge, J. Veldman, Joint Condition-based Maintenance and Load-sharing Optimization for Multi-unit Systems with Economic Dependency

2019008-EEF: Keller, J.T. G.H. Kuper, and M. Mulder, Competition under Regulation: Do Regulated Gas Transmission System Operators in Merged Markets Compete on Network Tariffs?

2019009-EEF: Hulshof, D. and M. Mulder, Renewable Energy Use as Environmental CSR Behavior and the Impact on Firm Profit

2019010-EEF: Boot, T., Confidence Regions for Averaging Estimators

2020001-OPERA: Foreest, N.D. van, and J. Wijngaard. On Proportionally Fair Solutions for the Divorced-Parents Problem

2020002-EEF: Niccodemi, G., R. Alessie, V. Angelini, J. Mierau, and T. Wansbeek. Refining Clustered Standard Errors with Few Clusters

2020003-I&O: Bogt, H. ter, Performance and other Accounting Information in the Public Sector: A Prominent Role in the Politicians' Control Tasks?

2020004-I&O: Fisch, C., M. Wyrwich, T.L. Nguyen, and J.H. Block, Historical Institutional Differences and Entrepreneurship: The Case of Socialist Legacy in Vietnam

2020005-I&O: Fritsch, M. and M. Wyrwich. Is Innovation (Increasingly) Concentrated in Large Cities? An Internatinal Comparison

2020006-GEM: Oosterhaven, J., Decomposing Economic Growth Decompositions.

2020007-I&O: Fritsch, M., M. Obschonka, F. Wahl, and M. Wyrwich. The Deep Imprint of Roman Sandals: Evidence of Long-lasting Effects of Roman Rule on Personality, Economic Performance, and Well-Being in Germany

2020008-EEF: Heijnen, P., On the Computation of Equilibrium in Discontinuous Economic Games

2020009-EEF: Romensen, G.J. and A.R. Soetevent, Improving Worker Productivity Through Tailored Performance Feedback: Field Experimental Evidence from Bus Drivers www.rug.nl/feb