MASTER'S THESIS

Learning in Complex Systems

June 2018

Performance Feedback:

A Literature Review and an Empirical Study

Aud Kielland Elnes

Faculty of Health Sciences

Department of Behavioral Sciences

OsloMet – Oslo Metropolitan University

Acknowledgements

I would like to thank everyone at Veidekke for trusting me with the opportunity to arrange this project. The department is kept anonymous to keep the participants so. I would like to thank all the drivers for your participation, time and patience, the Department Manager for being supportive and believing in the project, the Supervisor of the Transportation Unit and the Administrative Leader for all your help and support underway, and the Product Manager of Dynafleet at Volvo Trucks for teaching me the system and helping the project.

Thank you Sigridur Sigurjonsdottir and Kalliu Kouto for your patience, guidance and support, and to Elise Frølich Furrebøe for helping me start this project.

A big thank you to my supportive and understanding family, Tollef and Eva, for all your support and for tolerating the late nights working. My family in law, grandmother and friends who helped with babysitting and support. And to my father, who made this master thesis possible

Table of Content

Table of Tables

ARTICLE 1

ARTICLE 2

Table of Figures

ARTICLE 2

Abstract

Feedback is the most used independent variable in OBM research, it is cost effective, not intrusive and easy to implement. Article one presents a literature review on studies published in JOBM that investigates the effect of feedback. Applications of feedback are organized by feedback characteristics, and the consistency of effects, and compared to the results of the previous review. The results showed that feedback did not produce consistent effects and technological progress enables us to give feedback with new characteristics which calls for new categories of organization. Arguments were made that improvement is still needed regarding consistent use of the term feedback in JOBM. The studies in this review did not contain any feedback definitions, and many described the procedures accurately.

Article 2 presents an empirical study on the effect of using performance feedback alone and combined with instruction to reduce engine idling in 8 truck drivers. Baseline was followed by feedback contingent on performance, then instruction followed, then feedback was removed. During the intervention average group idling went from 19,3% at baseline to 14,1% in the last experimental phase. The idling continued to decrease in 8 of 9 drivers during second baseline to 11,7%. From baseline to the end of the observation period, average group idling decreased with 39,3%. A lot of the improvement was due to one individual who nearly halved his idling levels.

Keywords: Performance feedback, instruction, prompt, antecedent, task clarification, engine idling, Organizational Behavior Management, differential reinforcement, objective review, feedback characteristics, feedback only, organizational behavior management

MASTER'S THESIS

Learning in Complex Systems

June 201

Performance Feedback:

A Literature Review

Aud Kielland Elnes

Faculty of Health Sciences

Department of Behavioral Sciences

OsloMet – Oslo Metropolitan University

Abstract

Feedback is the most used independent variable in Organizational Behavior Management research, it is cost effective, not intrusive and easy to implement. The main purpose of this review was to investigate studies published in JOBM that showed the effect of feedback, using the categorization system of Balcazar, Hopkins, and Suarez (1985) and Alvero, Bucklin, and Austin (2001). The results showed that a) like in the previous review, feedback did not produce consistent effects and b) technological progress enables us to give feedback with new characteristics and that calls for new categories of organization and c) there is room for improvement regarding consistent use of terminology in JOBM articles investigating feedback as an independent variable. The implications of this is discussed.

Keywords: objective review, feedback characteristics, feedback only, organizational behavior management, group contingencies

Organizational Behavior Management

Organizational Behavior Management (OBM) is a subfield of behavior analysis and applied behavior analysis. It´s area of interest is the behavior of individuals and groups in organizational settings. OBM uses the theory from behavior science and principles and basic experimental research from applied behavior analysis, to create interventions with the purpose to improve behavior and efficiency in organizations, businesses, school, and community (Culig, Dickinson, McGee, & Austin, 2005). OBM developed into a separate field in late 60s and early 70s and became a well-established discipline during the 80s, and in this period the Journal of Organizational Behavior Management (JOBM) was established, publishing studies within the field (Dickinson, 2001). Many reviews have shown the effects of OBM interventions in businesses and human service settings (Dickinson, 2001). Culig et al. (2005) mentions that however successful, the field of OBM have been criticized by applied behavior analysis for years, for not relating practice to theory. OBM and behavior analysis is built on the theory of selection by consequences, put forward by Skinner (1953). Selection by consequences happens both on individual (individual selection) and on organizational (cultural selection) level. Observation of behavior is the only thing that decide if a consequence was punishing or reinforcing. Reinforcement is anything that increase the likelihood of a behavior to occur, and applied behavior analysis and OBM usually use reinforcing consequences to increase wanted behavior. Punishing consequences often produce unwanted side effects like counter control and is considered ethically problematic in planned interventions (Daniels & Bailey, 2014). In work life there are many ways in which aversive control in the form of punishing consequences are used, often unintended. Skinner (1953) argued that planned performance feedback could decrease the use of unplanned aversive control in organizations, based on examples from observations of behavior and verbal behavior of employees during feedback interventions.

Performance Feedback

Feedback was the most used independent variable in a review of the history of JOBM by VanStelle et al. (2012). In a recent review by Gravina et al. (2018) on OBM interventions in human services settings from 1990 to 2016, it was still the most consistently used independent variable. It is low cost, don't require expensive reinforcers, comprehensive training of personnel and is easy to implement (Prue & Fairbank, 1981). Performance feedback has been frequently used to improve teaching in academic settings, to increase healthy behavior, to improve customer service, to improve training situations, increase productivity, decrease absenteeism, and in behavior safety. Hattie and Timperley (2007) reviewed the impact of feedback in teaching and learning, Nicol, Thomson, and Breslin (2014) researched how feedback from peers evokes multiple learning behaviors; Moon and Oah (2013) investigated feedback effect on ergonomic seating positions; (Eikenhout & Austin, 2005) improved customer service using a package intervention; (Howard & Digennaro Reed, 2015) improved training of volunteers at an animal shelter; (Robinson & Dow, 2001) increased the service hours of social workers at a mental facility; (Matt C. Camden, Price, & Ludwig, 2011) investigated how feedback decreased absenteeism among health personnel; (Rantz & Houten, 2011) showed how feedback increased compliance with tasks that have few or none natural reinforcers, like instructions and safety routines.

Performance feedback is a familiar term in many different fields of study, (education, psychology, management and organizational behavior management) but not a technical term where there is broad agreement on its meaning (Duncan & Bruwelheide, 1985; Peterson, 1982). Different definitions of the meaning of feedback exists among the fields and within the them. A definition provided by Hattie and Timperley (2007, p. 102) within educational research, is "information provided by an agent (e.g., teacher, peer, book, parent, self,

experience) regarding aspects of one's performance or understanding… Feedback thus is a "consequence" of performance". In OBM, a feedback definition offered by Daniels and Bailey (2014, p. 157) is "to be considered performance feedback, information must serve at least two functions. First, it must tell you where you stand relative to some target or goal. Second, the performer must know what to improve." Mayer, Sulzer-Azaroff, and Wallace (2014, p. 483) defines feedback in this way: "in organizational behavior management, feedback generally is viewed as the return of information in a form that can influence behavior. "They describe the function of feedback "as a discriminative stimulus, signaling that reinforcers, or punisher, or something ambiguous, or nothing special is likely to follow that particular event". A behavioral analysis definition of feedback proposed by Mangiapanello and Hemmes (2015, p. 70) is "feedback is the name of an operant conditioning procedure in which there is presentation of an exteroceptive stimulus whose parameters vary as a function of parameters of antecedent responding."

Peterson (1982) criticized the field of OBM for giving "feedback" the status of reinforcement. He described feedback as "information on past performance" that can serve any number of functions since feedback is just a physical stimulus that come in many forms and can acquire many functions through conditioning. Therefore, he thought the discussion on which behavioral function it serves was irrelevant.

Prue and Fairbank (1981) encouraged the investigation of how the different characteristics of feedback (the source, the way it is conveyed, in which setting, how often it was given and the content) to better understand its behavioral functions. Duncan and Bruwelheide (1985) also proposed that feedback should be researched in terms of its different characteristics to see if characteristics of feedback correlated with behavioral functions. In this way a system of organization of feedback could evolve, making feedback interventions easier to design and analyze in terms of behavioral functions (Duncan & Bruwelheide, 1985,

p. 92). The reviews of Balcazar et al. (1985) and Alvero et al. (2001) organized studies with feedback as an independent variable, including package interventions, by combinations and characteristics and correlated the way feedback was implemented with the effect of feedback.

This paper is replication of the review by Alvero et al. (2001) with focus on studies that could report on the effect of feedback alone. The main purpose of this review is a) to investigate how recent studies have conducted research on the effect and functions of feedback, using the categorization system from the previous reviews to see if some characteristics seem to produce more consistent effect of feedback alone. b) The previous review reported on a decrease in the use of feedback, this review will compare the number of search results to the previous c) compare the results to those of the previous review and look for differences in characteristics and d) investigate how many articles include a definition of feedback and if they discuss the behavioral functions of feedback in their study.

This review differs from the method used in Alvero et al. (2001) on three points:

1) In the two mentioned reviews, they investigated 4 journals, Academy of Management Journal (AMJ), Journal of Applied Behavior Analysis (JABA), Journal of Applied Psychology (JAP), and Journal of Organizational Behavior Management (JOBM). This review focuses only on articles published in JOBM. The criteria from the previous review called for a lot of work with manual selection. Preliminary searches in the three remaining journals included numerous articles on analog studies or studies on feedback in special education settings. The latter can be natural learning environments for some individuals but have similarities to lab studies regarding the experimenter's knowledge about the individual and possibilities to limit other variables, versus more common workplace and organizational settings. In Alvero et al. (2001), 29 out of 43 studies are published in JOBM. For the scope of this paper, selecting journals only from JOBM seemed not ideal, but reasonable.

2) The two previous reviews investigated the characteristics of feedback and the different combinations feedback was implemented with. This review focuses on the interventions who in some way investigated the effect of feedback alone.

3) Alvero et al. (2001) investigated the timespan since the last review. This review is limited to a timespan of a little more than 5 years (2013 – May 2018).

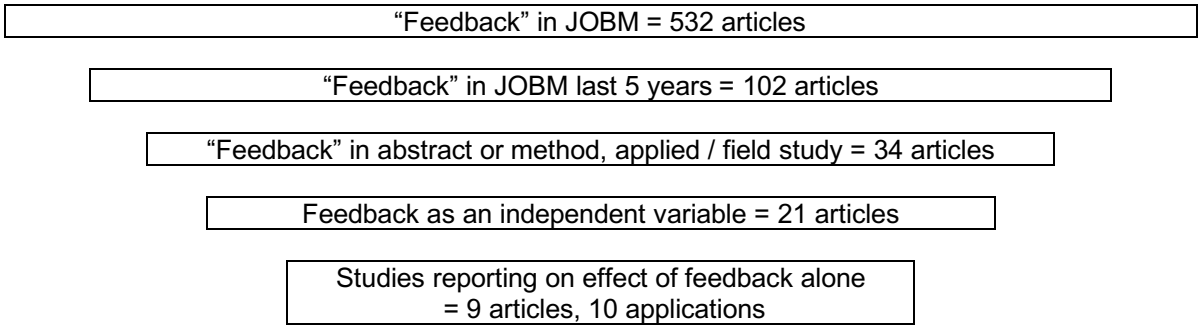
Method

The following article is a replication of Alvero et al. (2001), accurately describing the same criteria for categorization. Some elements are added, since some new characteristics occurred in the studies investigated in this review.

A search of the word "feedback" in JOBM gave a total of 532 results. The search period for this review was the last 5 years, January 1, 2013 to May 31, 2018, gave a total of 102 articles. As described in Alvero et al. (2001, p. 6), the articles were revised to exclude articles who did not have the word "feedback" in the abstract or method section, and articles who did not describe applied or field studies that targeted behavior in its natural environment. This excluded analog or laboratory studies. These criteria gave a total of 34 articles. As mentioned in the review by Alvero et al. (2001), articles describing feedback with other terms were not investigated, and that may have excluded some articles. The 34 articles were then revised to exclude those who didn't mention feedback as an independent variable. The remaining 21 articles were investigated to select studies who in some way investigated the effect of FB. This excluded studies that included FB as part of a larger intervention package, but included studies that implemented FB with other variables, as long as the design allowed for the study to investigate the effect of FB by either implementing FB before other variables or withdrew FB for part of the intervention. Continuing in line with the two previous reviews mentioned, each intervention that used feedback was classified as a separate "application" of

feedback (Alvero et al., 2001, p. 7). That gave a total of 10 interventions of feedback in the 9 articles revisited. Table 1 lists the articles included in this review, feedback applications found, and results of the variables.

Selection of Articles



Interobserver agreement

A second observer reviewed the articles left after the last step of the selection process described above. After discussion, three articles were removed; one study did not mention feedback as the name of the independent variable, one study was removed after a discussion on the definition of applied settings, and one study was removed after a closer inspection showed the intervention also included antecedents, and the intervention did not show the effect of feedback alone.

Both observers categorized the feedback characteristics according to the criteria listed below. Out of 10 applications and 7 criteria per application, the results were 11 disagreements out of 70, an agreement percentage of 84%. 6 of the disagreements were related to the category "content". After discussion, 2 new categories were added to the long list of content combinations. The categorization was copied directly from Alvero et al. (2001), but in retrospect an improved categorization system could have been made for this review. The long and unsystematic list of content combinations contributed to more confusion than overview. If one disregards the content – category the disagreements were 4 out of 70, a total of 92%

agreement. Disagreements were discussed until agreement was reached on the article selection and the categorization.

Data collection

Feedback applications were categorized and used as defined and described by Alvero et al. (2001), who replicated the categories from Balcazar et al. (1985). Each application was categorized according to effectiveness, feedback characteristics, and the feedback combination used. The following list of definitions will therefore be almost identical to the one in Alvero et al. (2001), except from two additional categories in "frequency" and "medium". The application of computerized feedback allows for feedback to be given "immediately following a behavior", and the feedback information being a "visual or auditory sign" that differs from the ordinary use of written and verbal feedback to such a degree that own categories were created.

Consistency of effects.

Feedback was categorized as "consistently effective" when it uniformly produced desired mean increases / decreases of performance compared with mean baseline levels and / or levels produced by any other independent variable(s). The effect had to be observed in all the participants, settings, and behaviors analyzed in order to be categorized as consistent. The effect of FB was categorized as "mixed" when it produced desired mean increases / decreases for some, but not all participants, settings, and behaviors analyzed in order to be categorized as consistent.

When mean performance levels during feedback was equal to mean baseline levels and / or when feedback effects were equivalent to the mean effects observed in a control group, they were categorized as "no effect". Studies where the baseline consisted of two data-points or fewer, did not allow for comparison of the effectiveness of the independent variables, and the effects were categorized as "unknown".

Feedback combinations.

Feedback was used alone or in combination with antecedents, goal setting procedures, and behavioral consequences. Studies where participants only received information about the quantity or quality of their performance were categorized as *"feedback alone"*. If the study referred to some form of antecedent stimuli, excluding goals (e.g., staff training, task analyses given to participants as information, job aids, weekly task objectives/assignments, supervisory prompts, etc.), it was categorized as "antecedent". "Behavioral consequences" could be events such as praise, monetary incentives, and time off work were given following desired changes in the target behaviors. "Goal setting" were a specification of a performance outcome or a standard of individual or group performance.

Feedback Characteristics.

The characteristics used to classify each feedback application are similar to those identified by Alvero. Minor changes, as described below, were made to two of the characteristics: source, medium (i.e., mechanism), frequency, and content.

Feedback Source. Feedback source referred to the individual or device that presented the information to the performer. Feedback sources were classified into: (a) supervisors and/or managers; (b) researchers; (c) self-generated feedback (from employees using a self-recording procedure); (d) customers; (e) mechanical devices; (f) experts (g) supervisors and researchers; (h) supervisors and self-generated feedback; and (i) studies that did not report the source of feedback.

Feedback Medium. Feedback medium were the means used to communicate the feedback information to the recipients: (a) graphs (display of individual and/or group performance); (b) verbal; (c) written; (d) verbal feedback and graphs; (e) verbal and written feedback; (f) verbal and written feedback and graphs; (g) written feedback and graphs; (h)

verbal and mechanical (e.g., videotape) feedback; and (i) visual / auditory signal (e.g. flashing lights or sounds.

Feedback Frequency. Feedback frequency referred to how often feedback was provided and was classified in nine intervals: (a) Continuous, after each occurrence, not specified how often per time unit. (following the behavior within a timespan of 60 seconds). (b) daily (one or more times in a period of 24 hours); (c) weekly (any frequency less than once per day and at least once per week); (d) monthly (any frequency less than once per week and at least once per month); (e) quarterly (any frequency less than once a month and at least once every four months); (f) daily and weekly; (g) daily, weekly and monthly; (h) weekly and monthly; (i) studies that did not report the frequency.

Feedback Participants. Feedback participants referred to whose performance was described by the feedback. The three types of participants were the following: (a) individual(s); (b) group(s); and (c) individuals and group(s) combined.

Feedback Privacy. Feedback privacy referred to how widely feedback information was made available: (a) publicly posted feedback (when feedback information was available not only to the performers, but also to other members of the organization); (b) private feedback (when feedback information was provided only to the performing individual); and (c) a combination of publicly posted feedback and private feedback (frequently used to separate group and individual performance or when some of the information was kept confidential).

Feedback Content. Feedback content identified the type of information provided. These content categories are also summarized in Table 2: A comparison of:
(a) an individual's performance with its previous performance; (b) a group's performance with its previous performance; (c) an individual's performance with a standard (e.g., a goal or a mean of performance) of individual performance (not information regarding previous

performance); (d) an individual's performance with his/her previous performance and a standard of individual performance; (e) a group's performance with a standard of group performance; (f) individual performance with group performance; (g) individual performance with a group goal; (h) individual performance with group performance and a standard of group performance; (i) group performance with a different group's performance; (j) individual performance, a standard of group performance, and a different group's performance; (k) individual and group performance with a standard of both individual and group performance; (l) information regarding a different group's performance; and (m) group performance with a standard and a goal; (n) a group's performance with its previous performance and a standard of group performance; (o) a sub-group to (the larger) group and a standard.

Results

Combinations

Alvero et al. (2001) found that the feedback alone - procedure was the most frequently used combination in that study (in 29% of the applications), and also in the previous review by Balcazar et al. (1985). This review contains only studies who in some way investigated the effect of feedback alone, logically leading to most applications in this review containing the feedback alone - procedure. Of 21 articles listing feedback as an independent variable, 9 of them (48%) used the "feedback alone" procedure. Alvero et al. (2001) found that the combination feedback and antecedents produced the most consistent results. In this review, the two applications that contained a combination procedure, combined feedback with antecedents and behavioral consequences.

Feedback characteristics

Table 3 show the distribution of feedback effectiveness as a function of the different feedback characteristics. None of the applications had no effect.

Source: In Alvero et al. (2001) the most frequently used source was supervisors, but the highest levels of consistent effects were correlated with supervisors and researchers. In this review the most frequently used source of feedback was supervisor (in three of ten applications), but only one of them produced consistent effects.

The sources related to most consistent results were researcher (100%), and mechanical (100%). Two of the studies did not clearly state who was giving the feedback. Only one application in Alvero et al. (2001) delivered feedback by a mechanical device, while in this study, 2 of 10 applications used a mechanical source.

Medium: In Alvero et al. (2001) the combination of graph with written or verbal feedback gave the highest level of consistency. In this review the most frequently used medium was verbal and graph in 4 of 10 applications, with 50% yielding consistent effect, and 50% mixed effects, and written feedback (three of ten applications) where 67% gave consistent effects. The medium that gave most consistent effects was Auditory signal, with 100% consistent effect.

Frequency: Alvero et al. (2001) concluded that weekly feedback was the most commonly used but least correlated with consistent effects. In this review weekly feedback was also the most commonly used (5 of 10), and the frequency with the least consistent effects. Continuous, daily and monthly was all correlated with 100% consistent effects.

Participants: In Alvero et al. (2001) individual was the most commonly used in applications, while group was correlated with the highest level of consistent effects.

In this review both group (5 of 10) and individual (4 of 10) were most commonly used, while the individual condition was correlated with the higher level of consistent effects (100%).

Privacy: In Alvero et al. (2001) the most commonly used was private, and the application correlated with highest levels of consistent effects was private and public feedback together. The most commonly used in this review was private (5 of 10) and public (4 of 10), applications of private produced the most consistent effects (80%) consistent while public gave 50% consistent effects.

Content: In Alvero et al. (2001) the most frequently used content was the a) individual performance compared to previous performance and c) individual performance compared to a standard of individual performance. The highest levels of consistency of effects were found in e) group performance compared with a standard, c) individual performance compared with a standard, and b) group performance compared with its previous performance. The most used content in this review was c) individual performance to a standard of performance, which also was correlated with 100% consistent effect. The second most used content was e) group performance to a standard of group performance (2 of 10) which gave 50% consistent effects.

Discussion

This review focused on the interventions who in some way investigated the effect of feedback alone. That selection criteria gave only feedback – alone applications and one study that investigated details within feedback frequency, So, Lee, and Oah (2013) investigated the effect of weekly vs. daily feedback.

Table 3 summarizes the main findings in the previous and present review. The small selection of articles in this review makes it hard to say that differences between them indicate

a trend in a certain direction. But some implications can be discussed, and some new characteristics were found.

Source: Supervisor was still the most frequently used source. In this review feedback from the supervisor did not produce consistent effects, while in Alvero et al. (2001) it was associated with high levels of consistent effects. The difference in effects among the two reviews could be related to this review concentrating on applications of feedback alone. Could feedback delivered by supervisor be perceived as punishment more often when the only intervention is feedback, than if it was implemented in combinations with other interventions? The sources related to most consistent results were researcher (100% and mechanical (100%). These are sources that might not be associated with, or in control of, punishing stimuli. Two of the 10 applications had mechanical source. Before the last step of selection, there were 2 additional articles using this source. Given the change in availability and price of technology the last years, it is likely that in the future, more feedback will be delivered by technology, which might cause a need for more specific or different categories. Two of the studies did not state who was giving the feedback, which is unfortunate, it is hard to investigate the behavioral function of feedback without details on how it is given.

Medium: In Alvero et al. (2001) the combination of graph with written or verbal feedback gave the highest level of consistency. The most frequently used medium in this review was verbal and graph in 4 of 10 applications, with 50% yielding consistent effect, and 50% mixed effects, and written feedback (three of ten applications) where 67% gave consistent effects. The most effective medium was Auditory signal, which gave a 100% consistent effect. Auditory or visual signals were not reported in the review by Alvero et al. (2001).

Frequency: In three of the applications, feedback was delivered continuous, immediately following behavior, and they all produced consistent results. Of the three with

consistent results there were a) one application of a loud "beeping" sound that conveyed information needed to fulfill the next step of the order picking process, and b) one application where the researchers sat in the room with the participants, giving them an auditory signal immediately following correct behavior and c) one application where the participants wore electrodes recording their sitting postures, and computerized feedback messages gave notice if the seating position was unhealthy. The feedback in these instances might function in case a) as a discriminative stimulus for completing a task and be reinforced by getting work done faster and in case b) as a conditioned reinforcer signaling task correctly executed, or the absence of the sound signaling the need for behavior change and c) as a discriminative stimulus for correcting seating posture and avoid getting another pop-up message on the computer. Feedback in the form of response – dependent delivery of short auditory or visual stimulus, like the brief sound of a clicker (Herron, Lotfizadeh, & Poling, 2018) or the short beep or flashing light of a machine (D. Goomas & Ludwig, 2017; D. T. Goomas, 2013), gave 100% of consistent effects (100%). That might mean an increase in visual and auditory signals, but also more nuanced feedback, as it takes less effort to give the feedback, and it can be delivered with closer proximity to the behavior of interest. In the study with the clicker (Herron et al., 2018) the researchers reported that although the participants rated the clicker feedback as more effective than verbal feedback given a while after behavior, the participants preferred the latter. This shows that not all that is proven effective is a good idea. For an employee it could be a stressful situation in the long run, to have someone observe them and giving immediate feedback continuously.

In their review, Alvero et al. (2001) noted that their search generated less articles than Balcazar et al. (1985) and questioned whether the use of the term feedback had gone down since the critique by (Peterson, 1982). Alvero et al. (2001) found 29 articles published in JOBM that studied feedback as an independent variable in applied settings, during a time

span of 14 years. This review covering only 5 years, found 21 articles that used feedback as an independent variable in applied settings in JOBM. Alvero et al. (2001) reviewed 43 articles, 14 of them used the feedback alone procedure in one or more applications. This review found 21 articles that filled the criteria set by Alvero et al. (2001). 8 of them (38%) used the feedback alone procedure. The numbers are not directly comparable since Alvero et al. (2001) investigated four journals, some of them could be more prone than others to use the feedback alone – procedure. But this indicates that the feedback alone - procedure is still widely used in studies published in JOBM, despite both the previous reviews concluding that feedback in combination with other procedures produced more consistent effects. After a possible decrease in the use of the term feedback and / or feedback interventions for a period after the critique from Peterson (1982) (Alvero et al., 2001, p. 22), the use of feedback interventions and the use of the term feedback, is perhaps increasing. This is also the conclusion from Houmanfar (2013, p. 86) who states that "Journal of Organizational Behavioral Management has seen reengagements of interest in performance feedback by researchers and practitioners over the years."

The critique from Peterson (1982) of the use of the term performance feedback was that there was not consistency in the terminology and definitions. Alvero et al. (2001) mentioned that the lack of consistency in the term feedback made it difficult to identify objective selection criteria that included all literature who used performance feedback as an independent variable. None of the articles selected in this review, although most of them used only feedback as an independent variable, mentioned a definition of feedback. The study by Clayton and Nesnidol (2017) is a good example of a study where missing definitions may lead to confusion. They wrote about feedback and prompt interchangeably, without defining the terms or tell the reader if it was one procedure or two in the method section.

The articles were reviewed to look for explanations of behavioral functions of their interventions. Some of the articles mentioned behavioral terms in the introduction or method, but not in discussion. Some mentioned a behavioral function in the discussion but did not really discuss their intervention. Others did discuss why their feedback intervention worked in terms of the function of feedback: Martinez-Onstott, Wilder, and Sigurdsson (2016) discussed feedback as consequence, D. Goomas and Ludwig (2017) mentioned the auditory feedback being a consequence of behavior, Matt C. Camden et al. (2011) discussed their intervention in relation to antecedents, behavioral consequence and reinforcement, Moon and Oah (2013) discussed their intervention in relation to negative reinforcement, So et al. (2013) mentioned reinforcement contingencies and antecedents as functions of their feedback intervention, Herron et al. (2018) commented on the function of their intervention in the title; conditioned reinforcers. Clayton and Nesnidol (2017) mentions conditioned reinforcement and antecedents, Pandey, Diller, and Miller (2016) wrote about antecedents and consequences, and D. Goomas and Ludwig (2017) mentioned reinforcement and interlocking contingencies. All of the articles mentioned behavioral functions related to their intervention.

Categorization: When more feedback is delivered by technology in the future, one could create categories under “medium” that separate between interventions where the signal follows a behavior, or when it also is antecedent for which behavior to perform next. There could be a category under “frequency.

As mentioned under inter observer agreement, a better categorization system is needed, to encompass the different versions of content, and a change in the frequency category is due after the occurrence of more studies with immediate and continuous feedback. There should be a category for proximity in time between feedback and behavior in addition to a frequency of feedback delivery. In that way one could differentiate between feedback given continuously, but in training sessions once a week for instance

Accuracy when describing interventions: When it comes to accuracy in describing feedback procedures, in the selection of articles for feedback alone – procedure it was discovered that many articles did not accurately describe that the feedback variable was part of an intervention package. Alvero et al. (2001) brought up this subject, urging accurate differentiation between feedback alone and package interventions.

In addition to the limited selection of articles, a limitation of this review was that it was not possible to have two researchers do the manual selection of all the 102 articles found in JOBM, and the rest of the selection process listed in the method section above.

This review found that feedback alone did not produce consistent effects. Two new categories that did produce consistent effects, were the mechanical medium and the continuous frequency. Many of the problems Alvero et al. (2001) is still widespread in the OBM literature. The lack of definitions in this selection of articles gave inconsistent use of terms in the studies, and inadequate descriptions of procedures made it hard to categorize the feedback characteristics in some reviews. Mangiapanello and Hemmes (2015) recognized problems with consistent terms in the literature and argued that behavior analysts should keep striving for conceptual consistency to avoid category mistakes.

References

Alvero, A. M., Bucklin, B. R., & Austin, J. (2001). An Objective Review of the Effectiveness and Essential Characteristics of Performance Feedback in Organizational Settings (1985-1998). *Journal of Organizational Behavior Management, 21*(1), 3-29. doi:10.1300/J075v21n01_02

Balcazar, F., Hopkins, B. L., & Suarez, Y. (1985). A critical, objective review of performance feedback. *Journal of Organizational Behavior Management, 7*(3-4), 65-89. doi:10.1080/J075v07n03_05

Camden, M. C., & Ludwig, T. D. (2013). Absenteeism in Health Care: Using Interlocking Behavioral Contingency Feedback to Increase Attendance With Certified Nursing Assistants. *Journal of Organizational Behavior Management, 33*(3), 165-184. doi:10.1080/01608061.2013.814521

Camden, M. C., Price, V. A., & Ludwig, T. D. (2011). Reducing Absenteeism and Rescheduling Among Grocery Store Employees With Point-Contingent Rewards. *Journal of Organizational Behavior Management, 31*(2), 140-149. doi:10.1080/01608061.2011.569194

Clayton, M., & Nesnidol, S. (2017). Reducing Electricity Use on Campus: The Use of Prompts, Feedback, and Goal Setting to Decrease Excessive Classroom Lighting. *Journal of Organizational Behavior Management, 37*(2), 196-206. doi:10.1080/01608061.2017.1325823

Culig, K. M., Dickinson, A. M., McGee, H. M., & Austin, J. (2005). An Objective Comparison of Applied Behavior Analysis and Organizational Behavior Management Research. *Journal of Organizational Behavior Management, 25*(1), 35-72. doi:10.1300/J075v25n01_02

Daniels, A. C., & Bailey, J. S. (2014). *Performance management : changing behavior that drives organizational effectiveness* (5th ed., rev. ed.). Atlanta, Ga: Performance Management Publications.

Dickinson, A. M. (2001). The Historical Roots of Organizational Behavior Management in the Private Sector. *Journal of Organizational Behavior Management, 20*(3-4), 9-58. doi:10.1300/J075v20n03_02

Duncan, P. K., & Bruwelheide, L. R. (1985). Feedback: Use and Possible Behavioral Functions. *Journal of Organizational Behavior Management, 7*(3-4), 91-114. doi:10.1300/J075v07n03_06

Eikenhout, N., & Austin, J. (2005). Using Goals, Feedback, Reinforcement, and a Performance Matrix to Improve Customer Service in a Large Department Store. *Journal of Organizational Behavior Management, 24*(3), 27-62.

Goomas, D., & Ludwig, T. D. (2017). Computerized Immediate Feedback Increases Product Recall Efficiency Due to Interlocking Contingencies in Food Manufacturing. *Journal of Organizational Behavior Management, 37*(1), 96-106. doi:10.1080/01608061.2016.1267067

Goomas, D. T. (2013). The Effects of Computerized Visual Feedback Using Flashing Lights on Order Selection in Large Industrial Settings: Productivity, Accuracy, and Cost Justifications. *Journal of Organizational Behavior Management, 33*(3), 209-220. doi:10.1080/01608061.2013.815098

Gravina, N., Villacorta, J., Albert, K., Clark, R., Curry, S., & Wilder, D. (2018). A Literature Review of Organizational Behavior Management Interventions in Human Service Settings from 1990 to 2016. *Journal of Organizational Behavior Management, 38*(2-3), 191-224. doi:10.1080/01608061.2018.1454872

Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational research, 77*(1), 81-112. doi:10.3102/003465430298487

Herron, M. A., Lotfizadeh, A. D., & Poling, A. (2018). Using Conditioned Reinforcers to Improve Behavior-Change Skills: Clicker Training for Practitioners. *Journal of Organizational Behavior Management, 38*(2-3), 172-190. doi:10.1080/01608061.2018.1454874

Houmanfar, R. (2013). Performance Feedback: From Component Analysis to Application. *Journal of Organizational Behavior Management, 33*(2), 85-88. doi:10.1080/01608061.2013.787002

Howard, V. J., & Digennaro Reed, F. D. (2015). An Evaluation of Training Procedures for Animal Shelter Volunteers. *Journal of Organizational Behavior Management, 35*(3-4), 296-320. doi:10.1080/01608061.2015.1093052

Lee, K., Shon, D., & Oah, S. (2014). The Relative Effects of Global and Specific Feedback on Safety Behaviors. *Journal of Organizational Behavior Management, 34*(1), 16-28. doi:10.1080/01608061.2013.878264

Mangiapanello, K., & Hemmes, N. (2015). An Analysis of Feedback from a Behavior Analytic Perspective. *An Official Journal of the Association for Behavior Analysis International, 38*(1), 51-75. doi:10.1007/s40614-014-0026-x

Martinez-Onstott, B., Wilder, D., & Sigurdsson, S. (2016). Identifying the Variables Contributing to At-Risk Performance: Initial Evaluation of the Performance Diagnostic Checklist–Safety (PDC-Safety). *Journal of Organizational Behavior Management, 36*(1), 80-93. doi:10.1080/01608061.2016.1152209

Mayer, G. R., Sulzer-Azaroff, B., & Wallace, M. (2014). *Behavior analysis for lasting change* (3rd ed. ed.). Cornwall-on-Hudson, N.Y: Sloan Publishing.

Moon, K., & Oah, S. (2013). A Comparison of the Effects of Feedback and Prompts on Safe Sitting Posture: Utilizing an Automated Observation and Feedback System. *Journal of Organizational Behavior Management, 33*(2), 152-162. doi:10.1080/01608061.2013.785906

Nicol, D., Thomson, A., & Breslin, C. (2014). Rethinking feedback practices in higher education: a peer review perspective. *Assessment & Evaluation in Higher Education, 39*(1), 102-122. doi:10.1080/02602938.2013.795518

NRK Metereologisk Institutt. (2018). Yr.no Blindern målestasjon. Retrieved from https://www.yr.no/sted/Norge/Oslo/Oslo/Oslo_(Blindern)_m%C3%A5lestasjon/statistikk.html

Pandey, N., Diller, J. W., & Miller, L. S. (2016). E-Mailed Prompts and Feedback Messages to Reduce Energy Consumption: Testing Mechanisms for Behavior Change by Employees at a Green University. *Journal of Organizational Behavior Management, 36*(4), 332-345. doi:10.1080/01608061.2016.1201034

Peterson, N. (1982). Feedback is not a new principle of behavior. *The Behavior Analyst*(5), 101–102.

Prue, D. M., & Fairbank, J. A. (1981). Performance feedback in organizational behavior management: A review. *Journal of Organizational Behavior Management, 3*(1), 1-16.

Rantz, W. G., & Houten, R. (2011). A Feedback Intervention to Increase Digital and Paper Checklist Performance in Technically Advanced Aircraft Simulation. *Journal of Applied Behavior Analysis, 44*(1), 145-150. doi:10.1901/jaba.2011.44-145

Robinson, K. E., & Dow, R. T. (2001). Using Graphical Feedback to Increase Social Workers' Productivity. *Administration in Social Work, 25*(3), 53-64. doi:10.1300/J147v25n03_04

Skinner, B. F. (1953). *Science and human behavior*. New

York: Macmillan.

So, Y., Lee, K., & Oah, S. (2013). Relative Effects of Daily Feedback and Weekly Feedback on Customer Service Behavior at a Gas Station. *Journal of Organizational Behavior Management, 33*(2), 137-151. doi:10.1080/01608061.2013.785898

VanStelle, S. E., Vicars, S. M., Harr, V., Miguel, C. F., Koerber, J. L., Kazbour, R., & Austin, J. (2012). The Publication History of the Journal of Organizational Behavior Management: An Objective Review and Analysis: 1998–2009. *Journal of Organizational Behavior Management, 32*(2), 93-123. doi:10.1080/01608061.2012.675864

Table 1. All Articles Included in This Review and the Results of All Variables.

Authors	Effects	Source	Medium	Frequency	Participants	Privacy	Content	Combinations
Matthew C. Camden and Ludwig (2013)	Mixed	Did not report.	Written & graphs	Weekly	Indiv. & group	Public & Private	(k) individual and group performance with a standard of both individual and group performance	FB alone
Clayton and Nesnidol (2017)	Consistent	Did not report.	Written (with illustration)	Weekly	Group	Public	(o) sub-group to large group and standard.	FB alone
D. Goomas and Ludwig (2017)	Consistent	Mechanical	Auditory signal	Continuous	Indiv.	Private	c) Ind. perf. to standard	FB alone
Herron et al. (2018)	Consistent	Researchers	Auditory signal	Continuous	Indiv.	Private	c) Ind. perf. to standard	FB alone
Lee, Shon, and Oah (2014)	Mixed	Supervisors	Verbal & graphs	Weekly	Group	Public	(n) Group to previous perf. and standard of group perf.	FB alone
Martinez-Onstott et al. (2016)	Consistent	Researchers	Verbal & graphs	Monthly	Indiv.	Private	c) Ind. perf. to standard	FB alone
Moon and Oah (2013)	Consistent	Mechanical	Written	Continuous	Indiv.	Private	c) Ind. perf. to standard	FB alone
Pandey et al. (2016)	Mixed	Experts	Written	Weekly	Group	Private	b) Group perf. to previous perf.	FB alone
So et al. (2013) *	Mixed	Supervisor	Verbal & graphs	Weekly	Group	Public	e) Group perf. to standard	FB, Ant & BC
	Consistent	Supervisor	Verbal & graphs	Daily	Group	Public	e) Group perf. to standard	FB, Ant & BC

Note. * This study investigated different details in the content – variable.

Table 2. Distribution of Feedback Effectiveness as a Function of Feedback Characteristics

	Feedback Effectiveness			
Feedback Characteristic	Consistent	Mixed	No	Total
SOURCE				
Supervisor/manager	1 (33%)	2	0	3
Researchers	2 (100%)	0	0	2
Expert	0	1	0	1
Mechanical	2 (100%)	0	0	2
Not reported	1 (100%)	1	0	2
	6	4	0	10
MEDIUM				
Written	2 (33%)	1	0	3
Verbal & graph	2 (50%)	2	0	4
Written & graph	0	1	0	1
Visual / auditory signal	2 (100%)	0	0	2
	6	4	0	10
FREQUENCY				
Continuous	3 (100%)	0	0	3
Daily	1 (100%)	0	0	1
Weekly	1 (20%)	4	0	5
Monthly	1 (100%)	0	0	1
PARTICIPANTS				
Individual	4 (100%)	0	0	4
Group	2 (20%)	3	0	5
Individual & group	0	1	0	1
	6	4	0	10
PRIVACY				
Private	4 (80%)	1	0	5
Public	2 (50%)	2	0	4
Private & public	0	1	0	1
	6	4	0	10
CONTENT				
(b) Group to prev. perf	0	1	0	1
(c) Individual perf. to standard	4 (100%)	0	0	4
(e) Group perf. to standard.	1 (50%)	1	0	2
(k) individual and group perf. to standard of individual and group perf.	0	1	0	1
(n) Group to previous perf. and standard of group performance	0	1	0	1
(o) Sub - group to group and standard	1 (100%)	0	0	1
	6	4	0	10

Note. Number and percentage of applications containing the different characteristics.

Table 3. Distribution of Feedback Effectiveness as a Function of Feedback Characteristics

	Alvero 2001		Present review	
Feedback	Most frequently used	Highest consistency effects	Most frequently used	Highest consistency effects
Combination	Feedback alone	Feedback & antecedents (100%)	Feedback alone*	Feedback alone*
Source	Supervisor / manager	Supervisor & researcher (100%)	Supervisor	Researcher (100%) Mechanical (100%) Not reported (100%)
Privacy	Private	Public & private (80%)	Private Public	Private (80%) Public (50%)
Participants	individual	Group (71%)	Group	Individual (100%)
Content	Individual Individual to standard individual	Group (71%) Group & stnd.group (75%)	Individual to stnd. ind. Group perf. to standard	Individual to stnd. ind. (100%) Sub - group to group and stnd. (100%) Group perf. to standard (50%)
Medium	Written	Written & graph (86%)	Verbal & graph Written	Visual / auditory signal (100%) Verbal & graph (50%)
Frequency	weekly	Daily & weekly (80%) Daily (71%) Monthly (80%)	Weekly Continuous	Continuous (100%) Daily (100%) Monthly (100%)

Note. * A comparison of combinations between Alvero et al. (2001) and the present review is not a fair comparison, as this review is limited to studies that is likely to have the feedback alone – combination.

MASTER'S THESIS

Learning in Complex Systems

June 2018

The Effect of Performance Feedback and Instruction

On Engine Idling:

an Empirical Study

Aud Kielland Elnes

Faculty of Health Sciences

Department of Behavioral Sciences

OsloMet – Oslo Metropolitan University

Abstract

This study examined the effect of using performance feedback alone and combined with instruction to reduce engine idling among 8 truck drivers at a contracting company. Using a within subject and within group design, baseline was followed by daily feedback contingent on individual and group performance, then instruction was added together with posting of group results and planning a social event not contingent on performance, before feedback was removed. During the intervention average group idling went from 19,3% at baseline to 14,1% in the last experimental phase. The idling continued to decrease in 8 of 9 drivers during second baseline to 11,7%. From baseline to the end of the observation period, average group idling decreased with 39,3%. A lot of the improvement was due to one individual who nearly halved his idling levels.

Keywords: performance feedback, instruction, prompt, antecedent, task clarification, engine idling, Organizational Behavior Management, differential reinforcement

Organizational Behavior Management (OBM)

Organizational Behavior Management (OBM) is a subfield of behavior analysis, building on the tradition of applied behavior analysis. The purpose of OBM is to improve performance and results in organizational settings, helping create good and effective workplaces by applying behavioral principles to the challenges of organizations. The behavioral principles OBM builds on is put forward by Skinner (1953): the selection of behavior by consequences on individual and cultural level. OBM help leaders identify contingencies in their organizations that fail to promote the wanted behavior or reinforces the wrong behaviors. It is common in OBM to design interventions based on reinforcement and to help organizations avoid side effects of aversive control(Prue & Fairbank, 1981). Many reviews have shown the effects of OBM interventions in businesses and human service settings (Dickinson, 2001). Among the most used strategies is performance feedback, cost effective and easy to implement because it doesn't require a lot of training. (Daniels & Bailey, 2014).

Performance feedback in OBM

Feedback was the most used independent variable in a review of the history of Journal of Organizational Management (JOBM) by VanStelle et al. (2012). In a recent review by Gravina et al. (2018) on OBM interventions in human services settings from 1990 to 2016, it was still the most consistently used independent variable. It is low cost, don't require expensive reinforcers, comprehensive training of personnel and is easy to implement (Prue & Fairbank, 1981). Reviews of the research on the effects of feedback found that it produces more consistent effects when combined with other procedure such as training, goal setting or praise than when implementing feedback alone (Alvero, Bucklin, & Austin, 2001; Balcazar, Hopkins, & Suarez, 1985). Feedback is often used in combinations with other elements like

goal setting or /and praise, and not much research is being done on the effects of feedback alone (Mangiapanello & Hemmes, 2015).

Despite disagreements on definition and behavioral functions of feedback, it remains an area of much interest to researchers and practitioners in OBM (Houmanfar, 2013). Several authors have pointed out that feedback seems to be the most cost effective intervention because it sometimes produces great changes with little effort (Daniels & Bailey, 2014; Prue & Fairbank, 1981). Johnson (2013) described how a feedback intervention that took less than one minute per worker to perform by the supervisor, produced a 17% - 30 % improvement on their average performance. Even if we don't have evidence of consistent effects (Alvero et al., 2001), it continues to be a popular intervention.

Daniels and Bailey (2014) argued that feedback should be considered socially important enough for the behavior analysis to investigate. Feedback applies to all fields of human performance and is seen as an essential part of learning. It is implemented in such an extended degree, being an important part of everyday life of students, employees and employers (Daniels & Bailey, 2014). Prue and Fairbank (1981) promoted the use of performance feedback in organizations despite inconsistent effects, because "it constitutes the most frequently employed behavioral strategy to change behavior in organizations", and that these kinds of planned interventions could take the place of unprogrammed aversive control procedures that often take place in organizations (Skinner 1953).

Performance feedback

What feedback actually is, the behavioral functions and definitions, are not any overall agreements on. Many different definitions exist, and there has long been a critique of the term feedback in behavior analysis, for being "professional slang" used to describe the properties of reinforcement, it takes the place of more correct behavioral terminology and

therefore should no longer be used (Peterson, 1982). Duncan and Bruwelheide (1985) argued that on the contrary, behavior analysis should become more engaged in defining the behavioral function of feedback and seek to influence the areas where the term is used. Kluger and DeNisi (1996) pointed out that since the effect of performance feedback is not always consistent, and it sometimes even produces negative effects, one should be careful when considering implementing feedback interventions at all, and not assume it will always give good results, although feedback interventions sometimes create positive effects. This points to an important area: Since definitions are rarely mentioned or the behavioral contingencies rarely described when feedback is implemented in applied research, the term feedback is often used inaccurately, which makes it hard to investigate why some feedback intervention works, and why others don't (Mangiapanello & Hemmes, 2015).

Prue and Fairbank (1981) defined feedback as "information given to the person about the quantity and quality of past achievements, and that in order to be effective, feedback should provide the necessary information for the individual to know what is the appropriate behavior, and it should classify characteristics about the rate of that behavior" (Prue & Fairbank, 1981, p. 9). They proposed that feedback takes on the functions of a discriminative stimulus in signaling reinforcement or negative consequences. Duncan and Bruwelheide (1985) wrote that "an operant perspective suggests that feedback is either a form of reinforcement or stimulus control, depending on the situation in which it occurs" (Duncan & Bruwelheide, 1985, p. 97). Mayer, Sulzer-Azaroff, and Wallace (2014) argued that feedback can function as a consequence (reinforcer or punisher) or have a discriminative function, and defined feedback as "information transmitted back to the responder following a particular performance in at form that may influence behavior: seeing or hearing about specific features of the results" (Mayer et al., 2014, p. 702). Alvero et al. (2001) wrote that although there is little consensus about what the term feedback contains, and little research has been done that

tests the features of the term, more research should be done in applied settings to investigate the significance of implementation (combinations and characteristics) on the effect of feedback. Alvero et al. (2001) wrote that feedback can act as an antecedent, reinforcer, establishing operation, and that it generates rule-directed behavior.

Daniels and Bailey (2014) argued that without an agreed upon definition of feedback, it is easily confused with "information", causing people to believe that they have given employees, patients or collogues what they need to change their performance, it just didn't work. Daniels and Bailey (2014) argued that we need to take into account what would actually count as feedback for the participants. Feedback is specific information that puts performance in context, and makes the individual aware of how they are doing, and what to change. Daniels and Bailey (2014) argues that feedback can take on many different functions, as a discriminative stimulus, a secondary reinforcer or punisher when conditioned by history of consequences. They defined performance feedback in this way: "To be considered performance feedback, information must serve at least two functions. First, it must tell you where you stand relative to some target or goal. Second, the performer must know what to improve." (Daniels & Bailey, 2014, p. 157).

Daniels and Bailey (2014); Mangiapanello and Hemmes (2015) argued that processes of operant conditioning explain the feedback – phenomena: "It may describe characteristics of the immediately prior response or of a predetermined target response (goal) and possibly the relation between the two … It may also describe the contingency between responding and the consequences of responding" (Mangiapanello & Hemmes, 2015, p. 54). They define feedback this way: "…feedback is the name of an operant conditioning procedure in which there is presentation of an exteroceptive stimulus whose parameters vary as a function of parameters of antecedent responding." Mangiapanello and Hemmes (2015, p. 70).

Daniels and Bailey (2014) focused on the recipients interpretation of the content; what makes feedback relevant to the recipient (they do write to teach the most efficient way to give feedback and get the wanted effect) while Mangiapanello and Hemmes (2015) covers the topic of the recipients "interpretation" in the description of feedback as an operant procedure. They also describe how feedback can contain an explanation of contingencies. If OBM starts using this definition it should cause a clarification in the terms prompt, task clarification, goal setting and so on.

Feedback is often shown more effective in combination with other procedures; goal setting, positive reinforcement, instructions or task clarification (Daniels & Bailey, 2014). Recent research has demonstrated that the combination of praise and specific information on performance is more effective than either presented alone, behavior must be reinforced to be maintained (Johnson, 2013). But the literature doesn't always explain or differentiate between feedback and other interventions it is combined with, and that makes it difficult to analyze which combinations are more effective (Alvero et al., 2001). Balcazar et al. (1985) and later (Alvero et al., 2001) investigated which elements of feedback is most important for producing consistent effects; the source (who gives it), the medium (how it is conveyed), the frequency (how often) and the content (what information), and added two elements; who the feedback was given to (individual / group / size of the group) and in which setting it was communicated (private or public), and the other elements it was combined with.

Feedback is considered to be most effective when the source is in control of the consequences, like the supervisor / manager (Daniels & Bailey, 2014). The combination of verbal or written feedback with graphs is said to be correlated with a high consistency of effects (Alvero et al., 2001). Research by Tosti (1987) argued that giving feedback immediately prior to behavior had a greater effect on quantity of behavior than if feedback came after the performance. Roberts and Rosales (1997) found that feedback given prior to

the next performance had a better effect on establishing new behavior, than feedback after performance. The more complex the task was, the better results were from feedback before task. Prue and Fairbank (1981) argued that considerations should be made when giving feedback, to avoid negative side – effects. Less fortunate performance data shared in public can lead to embarrassment or competition among collogues. This can drive behavior change, but if it causes the employees to perceive the intervention as aversive, it can create counter control and negative group contingencies that can have long term consequences. This also goes for deciding the content of the feedback – to be compared to others can be perceived as unfair if you don't have the same prerequisites for doing a good job.

Instruction

In the OBM literature, multiple words are used to describe a procedure that often precedes performance feedback; the performer receives information about how to perform the task ahead. Some words frequently used are instruction, information, prompt, job aids, task analysis, task clarification, teaching, telling and training. There are benefits to using a word that is recognizable for people who are not behavior analysts but might read research within OBM. On the quest to find a behavioral explanation to the word instruction, by The Oxford English Dictionary (Simpson, Weiner, & Press, 1989) defined as "a piece of information about a particular fact" and "knowledge or authoritative guidance imparted by one person to another». Daniels and Bailey (2014) listed instruction together with information as a subcategory of antecedents. Antecedents were defined as "circumstances, including signals and signs in our internal and external environment, which set the occasion for behavior" (Daniels & Bailey, 2014, p. 323).

Mayer et al. (2014) referred to the definition of *prompt* by (Touchette & Howard, 1984) saying that "prompts are stimuli that control the desired behavior but are not

functionally related to the task". Mayer et al. (2014) described *prompt* as something that encourages people to match their behavior to a standard of performance. They listed *instruction* as the same as *telling*, a subcategory of *prompt*; and defined telling as "an instructional or stimulus control procedure that uses oral, written, signed or other instructions or rules to prompt correct response under correct conditions, and enabling it to become eligible for reinforcement"(Mayer et al., 2014, p. 373), and offered a short everyday description of telling and instruction: "to sign, say or signal what to do".

For this paper the word *instruction* was used because it has a behavioral explanation as a subcategory of the terms *prompt* and *antecedent*, and at the same time it is close to the linguistic explanation of the term, and therefore intuitive for those who are not used to the behavioral terminology.

Rationale and Aims

Research is still needed on the different behavioral functions feedback can take on depending on how it is implemented (Mangiapanello & Hemmes, 2015). Even though feedback alone seem to not produce consistent effects (Balcazar et al., 1985), performance feedback is a very cost effective, not intrusive method, and therefore much used intervention both alone and in combination with other procedures like reinforcers, goal setting and praise. Intervention packages takes more resources, but produces more consistent effects (Daniels & Bailey, 2014).

Developing ways to provide more environmentally friendly services have become exceedingly relevant for the construction industry during recent years. Initiated by government regulations and an increased demand in the marked for environmentally responsible projects, requirements to win contracts now contains goals of low emission of nitrogen oxides (NOx, which causes pollution of the atmosphere (Simpson et al., 1989)

during construction processes. This study was done by adding feedback contingent on performance improvement, an information meeting on engine idling, and then withdrawal of feedback.

The aim of this study was twofold; to investigate the effect of performance feedback alone on engine idling, and to examine the effect of performance feedback combined with instructions on engine idling. The following research hypothesis were investigated:

H1: Performance feedback alone will lead to reduced idling.

H2: Feedback and instructions will have a larger effect on the target behavior than feedback alone.

H3: Feedback presented at the start of the workday will lead to reduced idling that day.

Method

Participants and Settings

Participants were eight professional truck drivers at Veidekke, a large contracting company. During normal workdays they drove trucks to different construction sites, collecting bulk fractions of building debris and transported it to different landfills.

Sampling Procedures

Before starting the experiment, the drivers were informed that the company wished to let a student carry out an experiment, and they were encouraged by the department manager to participate. The experimenter met with the daily manager (who coordinates the drivers workdays), explained the project, and had a separate information meeting with the drivers where they signed the consent form (see appendix A). They were informed verbally and in writing about the experiments aim to reduce NOX emissions by reducing idling time, and that the experimenter would send them text messages for the next 8 weeks contingent on their idling performance, and that they could opt out of the experiment at any time without giving

a reason, by texting the experimenter. Due to change of trucks in the middle of phase B, data could no longer be obtained for driver 4 and 5, and they were excluded from the study.

Feedback.

Frequency and scoring.

Feedback messages was sent contingent on a decrease in engine idling. The purpose was to a) differentially reinforce performance improvement and associate the messages with something positive and b) prevent habituation that could occur if feedback was given every workday for 8 weeks. Each morning the drivers who had improved in engine idling would receive feedback. If their performance yesterday was better than the average of the previous 4 days, they received a message. Scoring the performance compared to the previous 4 days instead of just yesterday, was intended to control for some of the variability in idling due to routes and traffic, preventing the feedback being given contingent on traffic pattern instead of the target behavior. If the group average improved, the group got a message, scored in the same way. The message to the group was sent at 7.15 o clock, and the individual message was sent 5 minutes later. Se examples of feedback messages below.

Content and source.

The individual feedback message contained information on individual performance compared to previous performance. The group message contained group results compared to the groups previous results, scored the same way. Sending a message on group performance in addition to individual performance, was an opportunity for the individuals to compare own results with the group, and to make sure that individuals who might not have gotten many messages, got messages on group performance. The content of the message was held as neutral as possible, not containing any intentional reinforcement such as praise, but fulfilling the definition of feedback from Daniels and Bailey (2014). The sender of the message was the researcher.

Template for text messages sent to drivers:

Individual message:
Idling time:
Your time Wednesday was 13 %
That is 2% less than the previous 4 working days.
Have a great day.

Group message:
The groups idling time:
Average time Wednesday was:16%
That is 1,2% better than group average the previous 4 working days.
Have a great day.

Instruction.

After the phase where feedback was given contingent on performance, there was arranged an information meeting with an instructor from Volvo Trucks. Below is an overview of the information given, which aligns with the definition of "telling" from Mayer et al. (2014): "instructional or stimulus control procedure that uses oral, written, signed or other instructions or rules to prompt correct response under correct conditions, and enabling it to become eligible for reinforcement". The meeting and social gathering was paid work hours, and there was food served. Preference assessment was made by asking the daily manager to suggest an activity his employees would like.

Information meeting and the contingencies described:

- Clarification of what is defined as idling.
- Why new engines make it less necessary to idle.
- Consequences of less engine idling: saving fuel, money, environment, company's reputation.
- Examples of consequences for the firm if idling was done too much in certaing situations.
- Posting of group results.

- Signaling a goal or standard of engine idling at 15% on group level being “normal”.
- Questions and discussion of situations where it is hard to avoid idling.
- Planning a social gathering as a thank you for participating in the study.

Experimental design.

Using an ABCA within-subject design, feedback and instruction was implemented on group level. Phase A: Baseline data was taken. Phase B: Feedback on idling performance was delivered daily, contingent on performance improvement. Phase C: A meeting with an instructor at Volvo Trucks containing information on idling. Group results were posted and there was signaled a positive consequence (social gathering) for participating in the study. In the second baseline phase, feedback was no longer delivered, and data was taken for 15 additional days.

A: Baseline 30 workdays.

B: Feedback contingent on performance improvement 19 workdays.

C: Information meeting and continuing feedback 11 workdays.

A: Second baseline 15 workdays.

Data Collection, Dependent Variable and Instrumentation.

The trucks were equipped with the Dynafleet management system (Volvo Trucks), designed for fleet management. It collects data from the cars tachograph and engine control unit on fuel use, pattern of driving, positioning and driving times. The dependent variable was engine idling. The Dynafleet system measures idling as time where the motor is running while the car is standing still, except from when the hydraulics of the car is activated, lifting the truck body when loading off debris, or operating a crane on the vehicle. The Dynafleet system measured engine idling as percentage of time during the day that the engine on while the truck is standing still. If the car was standing still with engine on while hydraulics was running (loading on or off), it did not measure as idling. Analysis was performed on a laptop

using Excel. Collection of data was accessed on Dynafleet web application Dynafleet Online (Volvo Trucks).

Feedback was distributed to the driver's cellphones via text-messages created in an online, paid text message service (Sveve.no) made for easy distribution of text messages for companies to groups and individuals. The messages were prepared in the program each evening and sent automatically next morning. The drivers and the daily manager reported that mobile phones were located on the dashboard of the cars while driving and used during the day to coordinate the different driving routes. This enhanced the chance of the drivers seeing the message soon after it was delivered. The information meeting was audio recorded and analyzed to check that it fulfilled the criteria of instruction.

Outliers.

Usually the working days consisted of 9-12 hours of driving. Driving times less than 120 minutes was very unusual and when these working hours occurred in the data collection, idling was often abnormally high, containing outliers such as 80-100% idling of total driving time. From 31. July 2017 to 12. April 2018, 23 data points were below 120 minutes and removed from the data collection: only 2 of the outliers occurred during the baseline and intervention.

Results

Insert Figure 1

Insert Table 1

Group Results

Average idling at baseline was 19,3%, with a variability of 3,4%.

The idling pattern decreased in phase B and in phase C and continued to decrease in second baseline. Average idling and variability also steadily decreased during the phases.

During the intervention idling average was reduced to 11,7% and variability to 2,8%.

Individual Results

Driver 1.

Average idling at baseline was 16,1%, with a variability of 6,1%.

The idling pattern in phase B and the start of C was similar to the high idling and high variability in baseline. A decrease in idling pattern started in middle of phase C, continued to decline during second baseline. During the intervention idling average reduced to 11,3% and a variability of 3,3%.

Driver 2.

Average idling at baseline was 10,9%, with a variability of 4,5%.

The idling pattern in second half of phase A was lower than in the first half. The idling pattern in phase B and C was similar to the second half of baseline, average idling and

variability decreased a little during phase B and stayed similar trhorugh phase C. There was a small decrease during second baseline in average idling and varaiblity.

During the intervention idling average reduced to 8,9% and a variability of 2,1%. Driver two was one of the three drivers with the lowest idling average at start of intervention.

Driver 3.

Average idling at baseline was 16,7%, with a variability of 5,1%.

The idling pattern in phase B was similar to baseline, with a small increase in average idling and variability. In phase C idling and variability decreased, and the decrease continued during second baseline.

During the intervention idling average reduced to 10,8% and a variability of 2,7%.

Driver 6.

Average idling at baseline was 18,5%, with a variability of 4,8%.

The idling pattern in phase B was lower than baseline, and average idling and variability decreased. In phase C and second baseline there are not enough data to conclude, the four data points gives an average idling of 10,1% and variailbity of 2,8, and the 3 datapoints in second baseline are much higher.

Driver 7.

Average idling at baseline was 51,2%, with a variability of 13,7%.

The idling pattern of phase B looked similar to baslien, and while average idling increased, varaibilty decreased notably. In phase C there was a markable shift in idling pattern, both the idling average and variability nearly halved. The idling pattern and average lasted through second baseline. During the intervention idling average reduced to 28,7% and a variability of 9,1%.

Driver 8.

Average idling at baseline was 9,8%, with a variability of 4,4%.

The idling pattern in second half of phase A was lower than in the first half. The idling pattern decreased in phase B, and that decrease lasted through phase C and second baseline. Average idling and variability decreased in each phase.

During the intervention idling average reduced to 3,3% and a variability of 1,3%. Driver eight was one of the three drivers with the lowest idling average at start of intervention.

Driver 9.

Average idling at baseline was 11,3%, with a variability of 3,6%.

The idling pattern in phase B was higher than baseline, with an increase in varaiblity. Idling pattern, average and variability increased in phase C. In second baseline a decline in idling pattern was seen for the first time during the intervention. Average idling decreased from 14,1% to 12,2%, and variability also decreased. The idling level ended at a slightly higher level than at baseline, 12,2% average and 4,7% variability. But the graph of idling pattern was decreasing, not stabilized. Driver nine was one of the three drivers with the lowest idling average at start of intervention.

Driver 10.

Average idling at baseline was 18,3%, with a variability of 5,7 %.

The idling pattern of phase B shows a decline in idling, the average and variability also cereased. The idling pattern continued in first part of phase C and then increased towards second baseline. Second baseline do not have enough data to conclude, phase C ended on an average of 12,4% average and 4,7% variability. Data points for second baseline is too few to conclude, but they were higher than in phase C.

Summary of Results

Insert Figure 2

Figure 2 show the average idling of each participant during the phases of the experiment. Group average idling was 19,3% at baseline, and was reduced to 17,4% in phase B, 14,1% in phase C and ended on 11,7% in baseline, a decrease of 19% percent from phase B to C on group level. From baseline to the end of the observation period, average group idling decreased with 39,3%.

The drivers that reduced their average idling visibly in phase B was driver 2, 6 and 8, and 10. Driver 1 also had a small decrease in idling. Driver 7 decreased variability considerably during phase B. During phase C, driver 7 and 10 reduced their idling further. Driver 9 increased average idling during phase C but had a decrease of 13,5% during second baseline, graph descending when measures stopped. During second baseline driver 1 and 3 decreased their idling notably.

To see if baseline was representative, historical data was consulted. Figure 3 show idling average of every 4 days for August 2017 to mid – December 2017, and during the experiment.

Insert Figure 3

For driver 1, 3, 6, 10 and 7, the idling values during baseline were higher during baseline than in the previous five months. This could be due to weather and temperature

conditions discussed below under "Variables that could affect the results". Looking at the results from the different phases in the intervention, the drivers reach a lower level of engine idling during the experiment, than during the previous five months. The implications of the artificially high baseline is taken into consideration when discussing the results of phase B.

Days With vs. Days Without Feedback.

Figure 4 showed that during phase B, 6 of the 8 drivers (driver 1, 2, 3, 8 and 10) had lower idling values on days started with feedback, than the days without feedback. In phase C, the pattern was there in the same drivers, plus driver 9, but the difference between days with and without feedback was much less evident, only notable in driver 1 and 3. Driver 7 had the biggest difference in idling between feedback and no feedback days in phase B, and in phase C, there was almost no difference between the two conditions.

Insert Figure 4

Discussion

There is support for hypothesis one, that feedback alone would lead to decreased engine idling. Engine idling decreased on group level and in some individuals, but not all. There was support for the second hypothesis, that feedback and instructions would have a larger effect on engine idling than feedback alone. The third hypothesis was only partially supported, that feedback presented at the start of the day would work as an antecedent and lead to reduced idling that day.

Hypotheses one: Performance Feedback Alone

During phase B the average idling percent of the group was reduced from 19,3% at baseline to an average of 17,4% in phase B. That is a decrease of 9,8% percent from one phase to the next on group level. Five drivers out of eight (2, 6, 8, 10 and 7) had a visible decrease in idling during phase B. But two of the baselines showed decreasing patterns before the intervention; could it be that a baseline during December and January on a traffic related behavior was not representative?

Historical data of the period prior to the experiment (figure 3, August 2017 to mid – December 2017) show that baseline was higher during baseline compared to the previous five months for drivers 1, 3, 6, 10 and 7. When results from phase B were compared to the historical baseline, it was evident that the improvement seen in driver 1 might just be a return to a previous normal level of idling. When compared to the historical baseline, the average idling values for driver 2, 6, 7, 8 and 10 proved to be lower during phase B than during any data points in the five month long historical baseline. Even if the effect varied across individuals, the results are considered to support the hypotheses that performance feedback alone lead to a decrease in idling on group level, and in five of the drivers. this is in line with the study by Alvero et al. (2001) and Balcazar et al. (1985), that feedback often does not produce consistent effects.

The drivers were sent on different routes and destinations every day. Possible negative effects of feedback were avoided in this phase by not disclosing to any of the drivers their performance compared to others, avoiding competition and other side effects that are useful when the individuals are in complete control of the behavior being measured (Mayer et al., 2014), but not a good idea if that part is unclear.

A stimulus in itself don’t change behavior unless it is a consequence or signals a consequence (Skinner, 1953). The behavioral functions of feedback during phase B was

initially planned to be an antecedent one, since feedback was given early in the workday, prior to behavior (Tosti, 1987). If it was an antecedent, what could it signal? Duncan and Bruwelheide (1985) wrote about how the source of feedback could be the deciding factor for feedback being perceived as a reinforcement or a stimulus control. Feedback in this experiment was delivered by a neutral person that they did not know, and without control of the consequences, but it could signal that they were being observed. If being observed was associated with a reinforcing or punishing consequence, the feedback likely served different functions for each driver, which is in line with the operational perspective of feedback argued by Mangiapanello and Hemmes (2015)

Hypotheses two: Performance Feedback with Instruction

Phase C is considered to have had an effect on five of the drivers, and a major effect on the one performer who influenced the group average more than the others.

Drivers 3, 6, 7, 8 and 10 had a notable decrease of idling during phase C. Driver 7 decreased average idling percentage by 45,7% during phase C. That counts for a lot of the groups average decrease in phase C. If one did not have historical data, or data of temperatures and weather, it would be easy to conclude that the idling levels correlated with warmer weather during April. But historical data (figure 3) show idling during second baseline reaching a far lower level than late summer and fall the year before.

The effect of phase C could also be concealed by some of the decrease in idling that happened in phase B, drivers that did not improve during phase C could have reached their potential and had little more to improve in phase C. Driver 7 decreased his idling notably more during phase C than phase B, and he had much higher idling values than any of the other drivers, which makes his impact bigger on the collective reduction in idling.

Phase C started with an information meeting. The content of the meeting is described above, and by the researcher categorized as instruction. Following the definition from Mayer

et al. (2014), something that encourages people to match their behavior to a standard of performance and tells the performer what to do, a stimulus control procedure that uses rules to prompt the correct response under correct conditions. The meeting also contained a posting of graphed group results from the past month and the drivers were told that there was room for improvement. By demand from the drivers, the instructor also told the average idling level he thought was the realistic goal (15% idling) based on improvements at other companies he worked with. Questions from the drivers were answered by the instructor, about idling and driving patterns. The instructor told examples of consequences for the firm if the idling was done at a high level in certain situations explaining the contingencies of the behavior of interest. The daily manager and department manager were present, and the reasons for focus on decreasing engine idling was repeated. This had been said before in a similar meeting prior to the experiment. At the end of the meeting a social event three week ahead of time was scheduled, not contingent on performance. That could have signaled a reinforcement, or a possible aversive stimulus – being presented with group results again. Many elements from the meeting could affected behavior, instruction on how, when and why to idle less, the managers being present, posting of group results, a standard of performance being presented, and the signaling of a social event. Daniels and Bailey (2014) argues that although package interventions make it hard to figure out what actually worked, since the elements used in OBM, like feedback, instruction and social gatherings, can take on different behavioral functions depending on the learning history of the individual, an intervention package will often work well when you can't research what would function as a reinforcer for each individual. Posting of group results comparing performance to others or a standard for performance could be aversive to some and perceived as punishment, or it could function as an antecedent to perform more in line with the rest of the group (Prue & Fairbank, 1981).

Did the Effects Last?

During the 11 workdays of second baseline, idling continued to decrease on group level, a decrease of 17% from phase C, similar to the 19% decrease between phase B and C. But this time there was not one driver taking credit for most of the decrease. Driver 1, 2, 3 and 7 continued to improve their idling. And driver 9 had a decrease for the first time during the observation period, a decrease of 13,5% from the previous phase, and a decrease in variability. That poses two questions a) what maintained the decrease in phase C and b) why did driver 9 suddenly improve?

Driver 9 suddenly improving could be a coincidence, maybe he got better routes with less traffic. Driver 9 was among the good performers to begin with. He had high variability, but his average idling was in the bottom three idling levels. Perhaps he did not have much to improve on. But, driver 2 had the lowest idling levels of all drivers and he still improved during the whole intervention. The social event was held in the middle of second baseline. On the group graph in figure 1, a “dip” in the graph can be seen, the graph seems to be rapidly decreasing towards the lowest point of idling. Two data points before that lowest point in the group idling graph, was the social event, where there was also a short posting of group results, and that seem like a valid explanation of the rapid decrease towards that point. The continuous decrease of idling after the intervention was over could also be due to other variables that that participants had little control over, and that is discussed below. The pattern of idling and the sudden decrease in group graph, does indicate that he social event either functioned as a reinforcer for creating good group results the final day of the project, or it could be an Sd for avoiding something aversive, display of bad results, the event could have created social contingencies that caused the drivers to talk about their results in a way that reinforced each other. The general continuous improvement could also be a result of the instructions containing different strategies to improve idling. Some were “quick-fixes” that

was easy to do right away. While rehearsing a new driving pattern to idle less in city traffic when approaching a traffic light or a que might take longer to get a hang of.

Both driver 1, 7 and 9 had decreasing graphs (figure 1) that seemed not to have found a point of stabilization yet at the end of observation. The study period ended before it could be seen whether the new levels of idling stabilized. Ideally, a longer second baseline should have been measured. If one had different control groups in a similar experiment, it would be interesting to investigate both the effect of instructions with feedback implemented on an off, and the effect of different instructions.

Some anecdotal observations were done by the researcher: at the start of the project there was a meeting where the drivers got information and signed a consent form. Of the questions they posed at the meeting, there was expressed a sense of unfairness from a couple of the drivers: why were *they* being measured – they already did a lot to improve environmental standard, they used the more expensive gas for their cars, and so on. On the information meeting there was observed both verbal behavior that indicated some though this was impossible to improve, while others asked to be shown the results at the social gathering. It seemed like there was some kind of change. Komaki, Heinzmann, and Lawson (1980) wrote about a similar incident, where the researcher observed a change in verbal behaviors in the workplace after group meetings that focused on safety levels at the workplace. They concluded that the meetings seemed to function as an occasion for the management, supervisors and employees to meet and talk about safety concerns in a positive way.

Hypotheses Three: Feedback as Antecedent for Less Idling

The last research hypothesis was that feedback present at the start of the workday would lead to reduced idling. Figure 4 show that 6 out of 8 drivers had lower idling percentage on days when feedback was delivered the same morning, but driver 6 and 9 had higher idling values on the days with feedback than without. The number of messages

received was investigated (appendix 2) to see if driver 6 and 9 received less messages than the other drivers. The drivers received 8 - 13 feedback – messages each, and driver 6 and 9 both received 11 messages, none of them received specifically more or less messages than the other drivers.

Perhaps some drivers had "more to work on" – those with the highest idling levels to begin with. In the study by Berger and Ludwig (2007), the feedback had the biggest effect on those who had made the most errors during baseline. Maybe feedback as an antecedent affected those who "needed it" the most. The results were investigated to see if driver 6 and 9 did worse than the others, and maybe didn't have much more to improve. Driver 9 was one of the bottom three on idling during baseline, so that explanation could be the case for him. But for driver 6 this theory seemed not to hold; he was among those who idled the most in baseline, and he also seemed to have an effect of phase B, since his idling decreased notably. Feedback did not seem to work as an antecedent for driver 6. Feedback presented prior to behavior seemed to contribute to the decrease in engine idling in some of the drivers. To work as an antecedent, feedback should be paired with a reinforcer to be effective. Daniels and Bailey (2014) suggested to pick the measure to reinforce so that you are sure to be able to reinforce regularly to avoid extinction. In phase B there was no obvious reinforcer except from the feedback itself, which did not contain any obvious praise, only information about how they had improved. But since the drivers were always compared to themselves in a four day average, the idling would improve sometimes – perhaps even without trying, they would get positive results and a message now and then. The feedback was given by the researcher not (yet) in control of any reinforcers. This indicates that for some of the drivers, information on their performance and how much it had improved functioned as a reinforcer, perhaps because of the design of pairing receiving feedback with it being a notion on improvement,

and that it happened often enough. The other possible explanation could be the observer effect, being observed could work as a discriminative stimulus for possible consequences.

Variables that Could Affect the Study

Variability.

The drivers had high variabilities in routes, often changing routes several times during one workday, which together with variations in traffic likely contributed to the high variability in idling. This can make it difficult to see whether the pattern and variation in idling values during the intervention was due to driving pattern or "good / bad" route or traffic. Looking at the results for group average idling may be a better way to see the effect, as it controls for this variation within each driver and between the drivers. Assignment of different routes on a regular basis creates randomization that could strengthen the external validity, as every driver would get both "good and bad" routes over the course of the intervention.

Weather, Temperatures and Representativeness of Baseline.

Weather and temperature can affect traffic and engine idling. There was an increase in engine idling December and January (see figure 3) during part of the intervention´s baseline. To control for this during the experiment, temperature data was collected and a correlation analysis using Pearson R correlation coefficient was done on the correlation between daily temperature and idling during the observation period. The correlation for the experimental period showed that there was a low level of correlation between temperature and idling during baseline and intervention (see appendix D).

Snow could have affected traffic. A graph from NRK Metereologisk Institutt (2018) shows snow coming in December, and more heavy snowfall starting mid-January continuing through February caused 40-55 cm snow to stay until the beginning of April (see appendix E). That means there was a fair amount of snow on the roads as phase B started, and all the

way through phase C. The snow disappeared from the roads about two weeks before phase C was finite. The idling went down in some of the drivers during phase B despite snowfall in the same period, but the snow could have prevented some of the effect of phase B. The snow disappearing a few weeks before end of phase C could have contributed to a decrease in idling during phase C. Two things indicate that the snow was not a major factor for the decline in idling: a) During the intervention the idling values decreased to levels below those of August last year and b) The snow disappeared in the middle of phase C, and the decrease in idling continued for several drivers through second baseline.

Reliability in measurements.

All data was collected from the Dynafleet system, which made the data collection vulnerable to technical errors. This was controlled for by also collecting data on the use of hydraulics. If the "hydraulic" measurement went unnaturally low, there was likely something wrong with the measurement of idling, since the two were connected. All the cars showed normal measures in the use of hydraulics, and the use of hydraulics did not decrease from normal levels during the intervention.

Conclusions

In this study the source was kept as neutral as possible, delivering performance feedback contingent on improvement, aiming to differentially reinforce behavior, then adding instructions and a social reinforcement for participating, but not for improving performance.

Performance feedback alone correlated with a visible decrease in idling in several drivers during phase B, continued to decrease during phase C and the decrease lasted during second baseline, still decreasing at the end of the observation period. The results were compared to historical data because of a high baseline, and this comparison showed that several drivers reduced their idling notably to below levels last seen around summer time,

some even during phase B. Correlation analysis also showed low correlation between idling and temperatures during the observation period.

The behavioral function of feedback in this experiment was planned to be an antecedent for behavior the same day. Although the numbers did show a pattern: most drivers had lower idling on days starting with feedback, the difference varied from 2 to 5 percentage points, and it is hard to know with limited time, how much of this was a coincidence.

The combination of feedback and instruction could have functioned as a reinforcer, or an antecedent for reinforcement or punishment. It is however hard to identify which elements had the most effect on behavior change in phase C. The continuing decrease in idling through second baseline could be an argument for instruction on different driving patterns causing behavior change, but also for the social event being an antecedent for behavior. Feedback could have functioned as a reinforcer for some individuals, or the experience of being observed functioned as an antecedent for reinforcement or punishment. In package interventions different combinations will work for different people at different times. Some elements might work as antecedents or reinforcers to some, while the same elements can have different behavioral functions to others, it depends on learning history what is perceived as reinforcing, or what signals reinforcement. (Daniels & Bailey, 2014), perhaps this is why intervention packages seem to produce more consistent effects in OBM studies.

This study was a hybrid of an experiment and practical problem solving. The results from this study indicate that the intervention created effects that lasted, that feedback alone did affect idling and that feedback combined with instruction, posting of results and signaling a social event, had more impact than feedback alone on idling behavior.

The effect of the interventions was not consistent: not all subjects decreased their idling in phase B or phase C. It is hard to know just how essential feedback was in the process without a control group that only received instruction. To better investigate the

behavioral functions of feedback, a similar study should have a control group. Perhaps instruction should have been the first variable implemented in a reversal design in one group, where instruction is given prior, and feedback is applied, removed and applied again. Kluger and DeNisi (1996) argued that a lot of applied research investigates whether feedback improves performance, but not the details of the processes involved in feedback effects. This experiment shows the dilemma of doing applied studies of feedback and attempt to investigate the behavioral functions at the same time.

References

Alvero, A. M., Bucklin, B. R., & Austin, J. (2001). An Objective Review of the Effectiveness and Essential Characteristics of Performance Feedback in Organizational Settings (1985-1998). *Journal of Organizational Behavior Management, 21*(1), 3-29. doi:10.1300/J075v21n01_02

Balcazar, F., Hopkins, B. L., & Suarez, Y. (1985). A critical, objective review of performance feedback. *Journal of Organizational Behavior Management, 7*(3-4), 65-89. doi:10.1080/J075v07n03_05

Berger, S. M., & Ludwig, T. D. (2007). Reducing Warehouse Employee Errors Using Voice-Assisted Technology That Provided Immediate Feedback. *Journal of Organizational Behavior Management, 27*(1), 1-31. doi:10.1300/J075v27n01_01

Daniels, A. C., & Bailey, J. S. (2014). *Performance management : changing behavior that drives organizational effectiveness* (5th ed., rev. ed.). Atlanta, Ga: Performance Management Publications.

Dickinson, A. M. (2001). The Historical Roots of Organizational Behavior Management in the Private Sector. *Journal of Organizational Behavior Management, 20*(3-4), 9-58. doi:10.1300/J075v20n03_02

Duncan, P. K., & Bruwelheide, L. R. (1985). Feedback: Use and Possible Behavioral Functions. *Journal of Organizational Behavior Management, 7*(3-4), 91-114. doi:10.1300/J075v07n03_06

Gravina, N., Villacorta, J., Albert, K., Clark, R., Curry, S., & Wilder, D. (2018). A Literature Review of Organizational Behavior Management Interventions in Human Service Settings from 1990 to 2016. *Journal of Organizational Behavior Management, 38*(2-3), 191-224. doi:10.1080/01608061.2018.1454872

Houmanfar, R. (2013). Performance Feedback: From Component Analysis to Application. *Journal of Organizational Behavior Management, 33*(2), 85-88. doi:10.1080/01608061.2013.787002

Johnson, D. A. (2013). A Component Analysis of the Impact of Evaluative and Objective Feedback on Performance. *Journal of Organizational Behavior Management, 33*(2), 89-103.

Kluger, A. N., & DeNisi, A. (1996). The effects of feedback interventions on performance: A historical review, a meta-analysis, and a preliminary feedback intervention theory. *Psychological Bulletin, 119*(2), 254-284. doi:10.1037/0033-2909.119.2.254

Komaki, J., Heinzmann, A. T., & Lawson, L. (1980). Effect of training and feedback: Component analysis of a behavioral safety program. *Journal of Applied Psychology, 65*(3), 261-270. doi:10.1037/0021-9010.65.3.261

Mangiapanello, K., & Hemmes, N. (2015). An Analysis of Feedback from a Behavior Analytic Perspective. *An Official Journal of the Association for Behavior Analysis International, 38*(1), 51-75. doi:10.1007/s40614-014-0026-x

Mayer, G. R., Sulzer-Azaroff, B., & Wallace, M. (2014). *Behavior analysis for lasting change* (3rd ed. ed.). Cornwall-on-Hudson, N.Y: Sloan Publishing.

NRK Metereologisk Institutt. (2018). Yr.no Blindern målestasjon. Retrieved from https://www.yr.no/sted/Norge/Oslo/Oslo/Oslo_(Blindern)_m%C3%A5lestasjon/statistikk.html

Peterson, N. (1982). Feedback is not a new principle of behavior. *The Behavior Analyst*(5), 101–102.

Prue, D. M., & Fairbank, J. A. (1981). Performance feedback in organizational behavior management: A review. *Journal of Organizational Behavior Management, 3*(1), 1-16.

Roberts, P., & Rosales, J. (1997). Should corrective feedback come before or after responding to establish a "new" behavior? In: ProQuest Dissertations Publishing.

Simpson, J. A., Weiner, E. S. C., & Press, O. U. (Eds.). (1989). Oxford: Clarendon Press.

Skinner, B. F. (1953). *Science and human behavior*. New

York: Macmillan.

Sveve.no. Retrieved from https://www.sveve.no/SMS/SendSMS

Tosti, D. T. (1987). Formative feedback. *Performance & Instruction, 26*(2), 18-21. doi:10.1002/pfi.4160260209

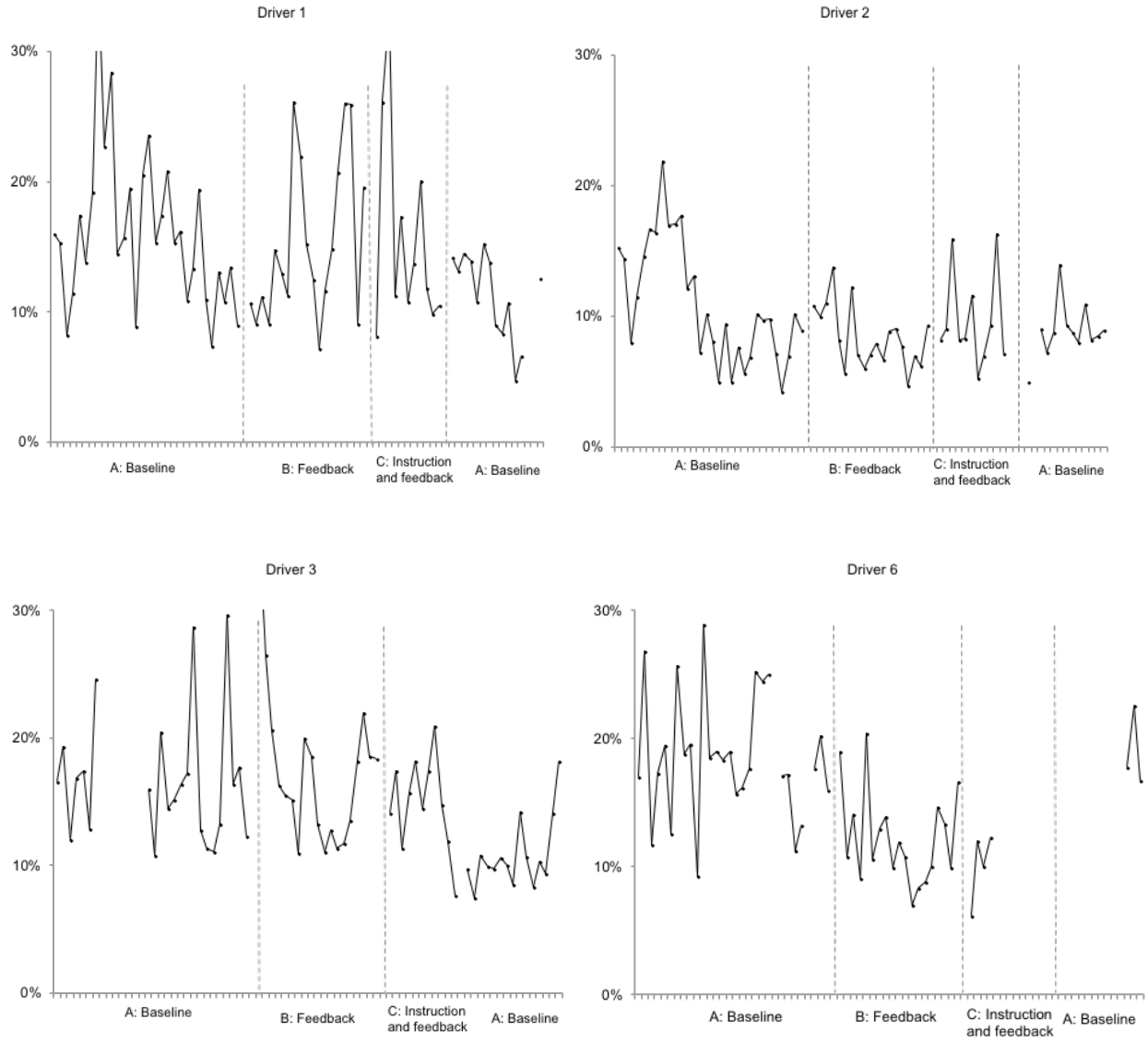
Touchette, & Howard. (1984). ERRORLESS LEARNING: REINFORCEMENT CONTINGENCIES AND STIMULUS CONTROL TRANSFER IN DELAYED PROMPTING. *Journal of Applied Behavior Analysis, 17*(2), 175-188. doi:doi:10.1901/jaba.1984.17-175

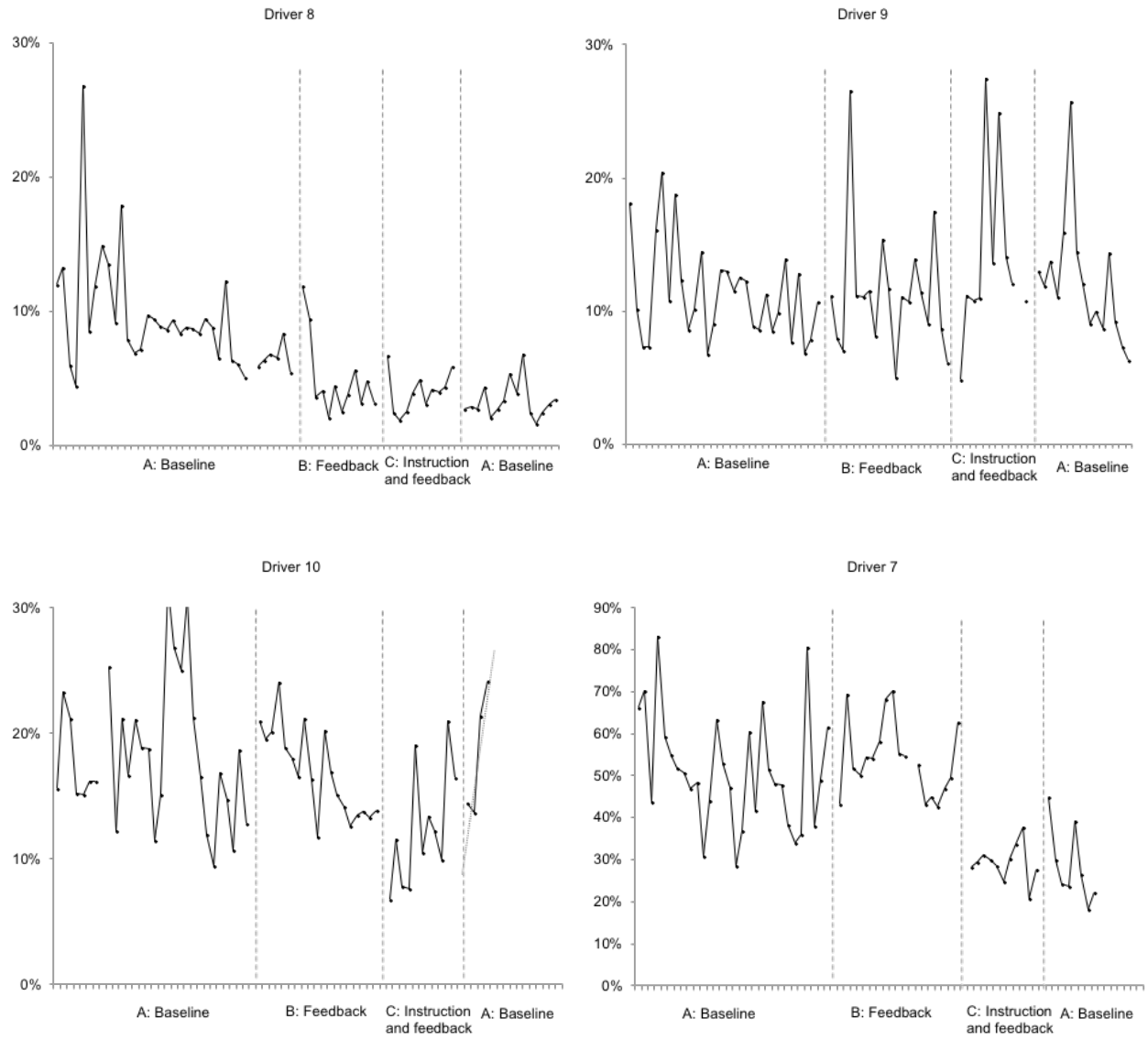
VanStelle, S. E., Vicars, S. M., Harr, V., Miguel, C. F., Koerber, J. L., Kazbour, R., & Austin, J. (2012). The Publication History of the Journal of Organizational Behavior Management: An Objective Review and Analysis: 1998–2009. *Journal of Organizational Behavior Management, 32*(2), 93-123. doi:10.1080/01608061.2012.675864

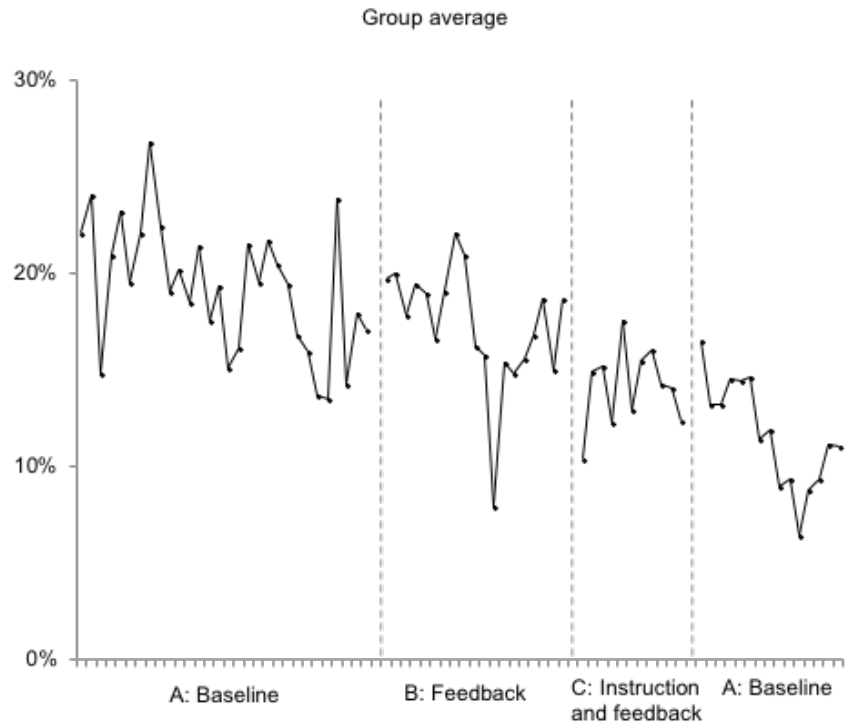
Volvo Trucks. (2018a). Fleet Management System - Dynafleet Retrieved from http://www.volvotrucks.co.uk/en-gb/services/dynafleet.html

Volvo Trucks. (2018b). www.dynafleetonline.com [Web application]. Retrieved from https://www.dynafleetonline.com/AS/authorize?response_type=code&client_id=dfol&redirect_uri=https://www.dynafleetonline.com/fm/portal/app/authenticate&scope=dfol-fm&state=anticsrf1004856070039004842

Figure 1: Percentage of Idling During the Experiment.

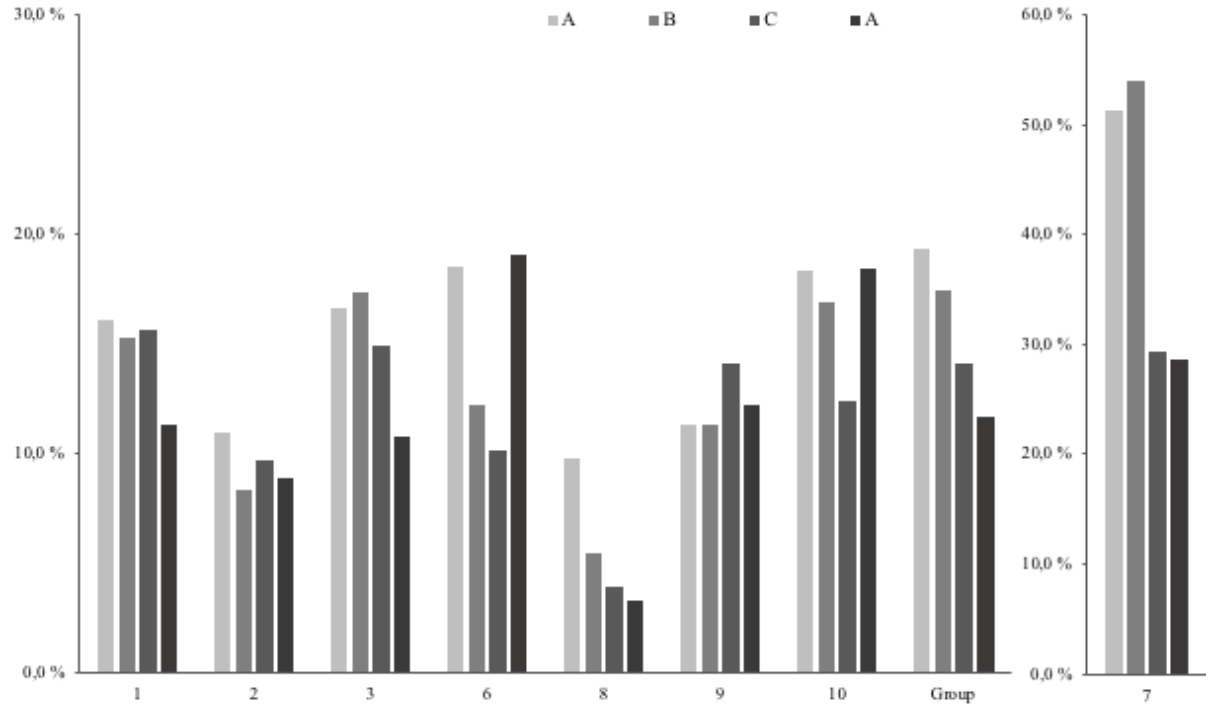



Driver 8
30%
20%
10%
0%
A: Baseline
B: Feedback
C: Instruction and feedback
A: Baseline
Driver 9
30%
20%
10%
0%
A: Baseline
B: Feedback
C: Instruction and feedback
A: Baseline
Driver 10
30%
20%
10%
0%
A: Baseline
B: Feedback
C: Instruction and feedback
A: Baseline
Driver 7
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
A: Baseline
B: Feedback
C: Instruction and feedback
A: Baseline



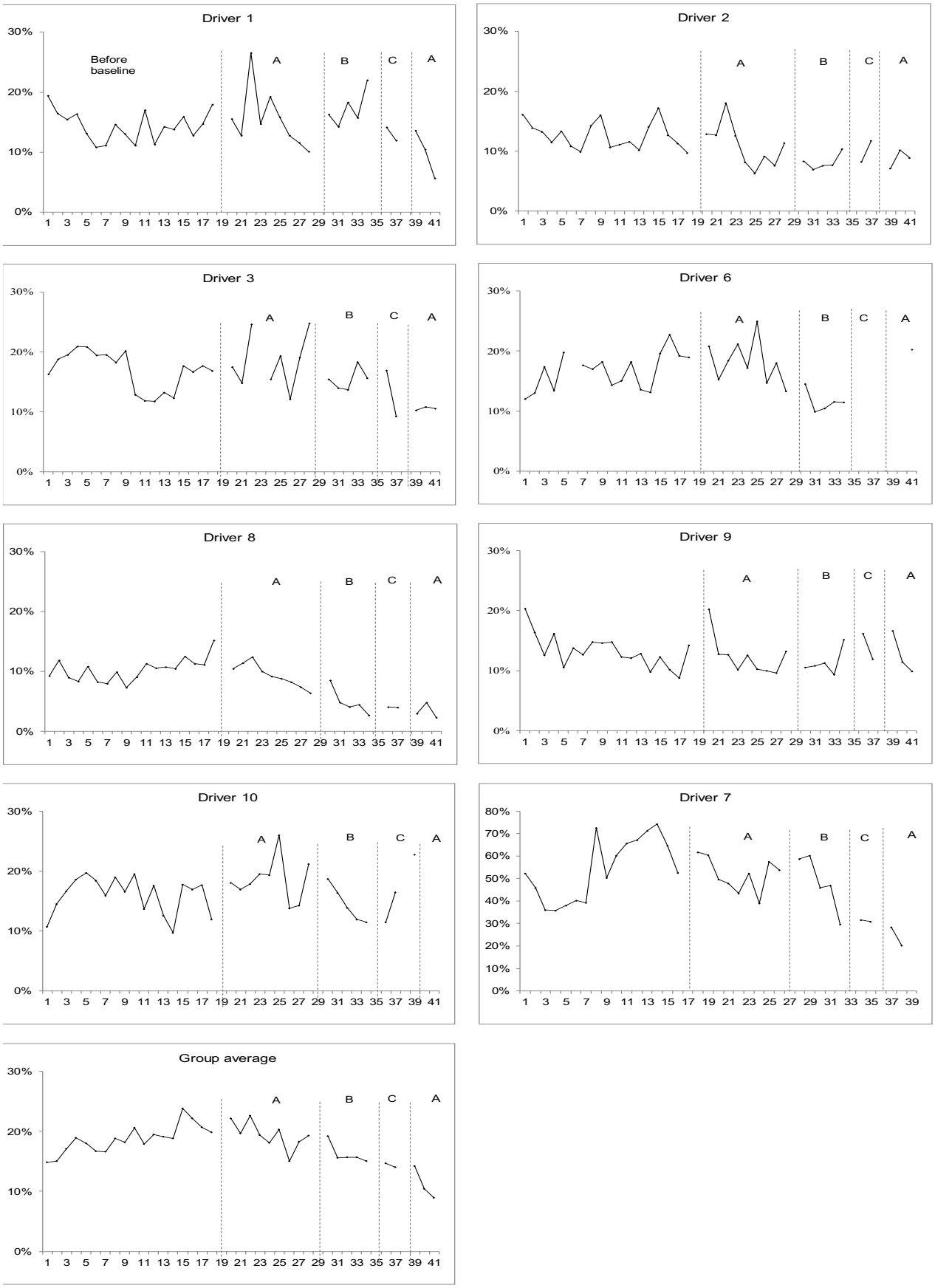
Note. The Y-axes is the average idling time of total driving time each day. Gaps in data is due to drivers having the day off. Driver 7 have different values on the y-axes than the rest, because of higher idling values.

Figure 2: Average Idling in the Experimental Phases.



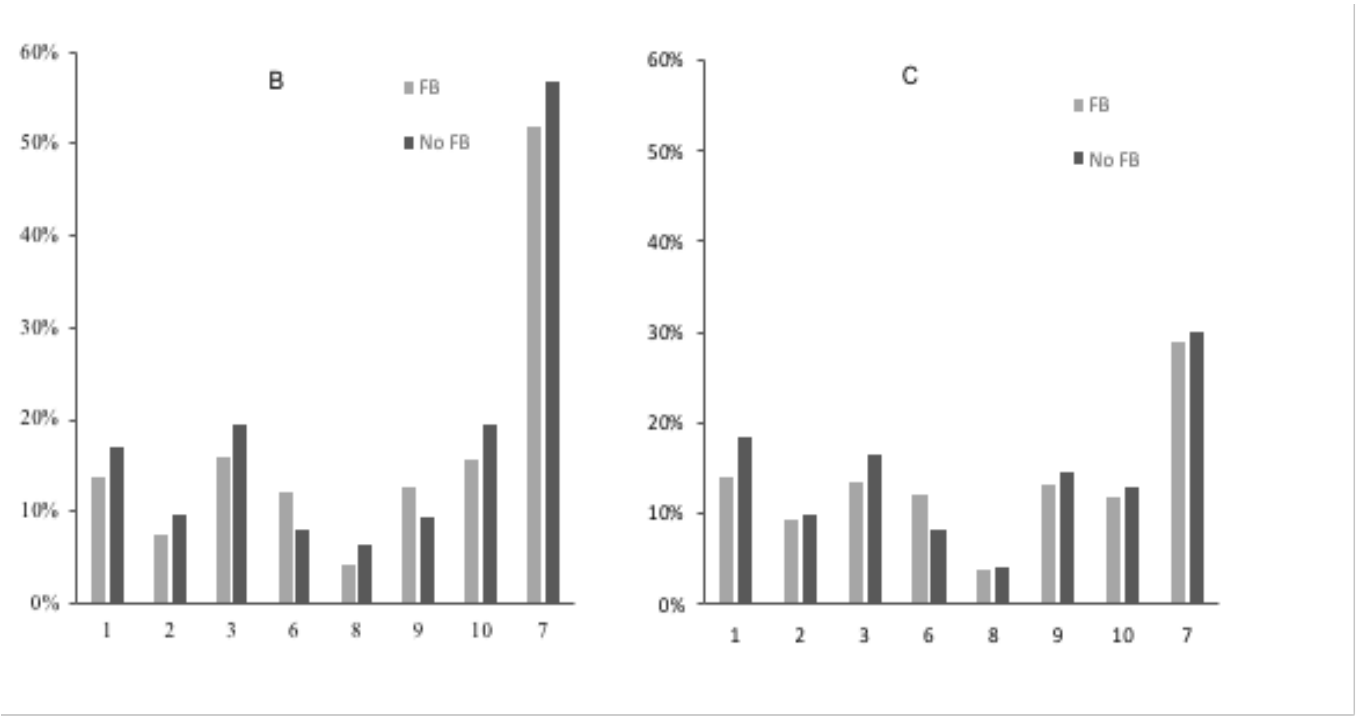
Note. The Y-axes is the average idling time of total driving time each day. Gaps in data is due to drivers having the day off. Driver 7 have different values on the y-axes than the rest, because of higher idling values.
*Driver 6 and 10: Too few data points in second baseline makes that bar not representative.

Figure 3: Historical Data



Note. Idling 4 days average, from August 2017 to May 2018. Baseline for the intervention was mid-December 2017 to mid – February 2018. The intervention took place in mid - February – April 2018.

Figure 4: Average Idling on Days with and Without Feedback, Phase B



Note. The Y-axes is the average idling time of total driving time each day. Driver 7 have different values on the y-axes than the rest, because of higher idling values.
*Driver 6 and 10: Too few data points in second baseline makes that bar not representative.

.

Table 1. Results of The Experiment in Different Phases.

Driver	Parameters	Baseline	Phase B	Phase C	2. Baseline
1	Average	16,1 %	15,3 %	15,7 %	11,3 %
	Variability	6,1 %	6,2 %	7,8 %	3,3 %
2	Average	10,9 %	8,4 %	9,7 %	8,9 %
	Variability	4,5 %	2,4 %	3,5 %	2,1 %
3	Average	16,7 %	17,4 %	14,9 %	10,8 %
	Variability	5,1 %	6,1 %	3,7 %	2,7 %
6	Average	18,5 %	12,2 %	10,1 %*	19,0 %*
	Variability	4,8 %	3,6 %	2,8 %*	3,1 %*
7	Average	51,2 %	54,1 %	29,4 %	28,7 %
	Variability	13,7 %	8,9 %	4,3 %	9,1 %
8	Average	9,8 %	5,4 %	4,0 %	3,3 %
	Variability	4,4 %	2,5 %	1,5 %	1,3 %
9	Average	11,3 %	11,3 %	14,1 %	12,2 %
	Variability	3,6 %	4,8 %	6,9 %	4,7 %
10	Average	18,3 %	16,9 %	12,4 %	18,5 %
	Variability	5,7 %	3,5 %	4,7 %	5,2 %
Group	Average	19,3 %	17,4 %	14,1 %	11,7 %
	Variability	3,4 %	3,1 %	2,0 %	2,8 %

Note. Variability is calculated by standard deviation. Grey areas marked are values that have decreased compared to the previous phase.
*Driver 6 and 10: Too few data points in second baseline makes those values not valid.

Appendix A. Consent Form.

SAMTYKKESKJEMA

Forespørsel om deltakelse i forskningsprosjekt
Bakgrunn og formål
Veidekke ønsker å redusere utslipp av NOx - gass og partikler for å bli bedre i stand til å vinne anbud hvor lave utslipp er kriterier. Dette prosjektet er en del av min masteroppgave ved institutt for atferdsvitenskap på Oslo Metropolitan University (tidligere HiOA). Mitt institutt forsker på ulike mekanismer for læring og valgatferd. Forespørselen om deltakelse i studien går til sjåførene ved Avd. for Transport * i Veidekke. Jeg ønsker å gjøre et eksperiment som kan være morsomt for deltakerne som er med, med mulighet for å generere positive resultater for avdelingen, i tillegg til å være et interessant forskningsprosjekt for masteren.

Hva innebærer deltakelse i studien?
Jeg ønsker å

- Sende deg en sms i løpet av arbeidsdagen (man – tors) i ca. 6-8 uker. Du blir spurt om å se på meldingen så fort du får anledning i løpet av arbeidsdagen, men ikke å svare på den.
- Invitere deg til et informasjonsmøte om tomgangskjøring hos Volvo, med Jan Erik Pedersen fra Volvo Norge, og Avd. leder i Veidekke, Dag Kristian Storhaug.

Data som registreres vil være data som allerede samles inn i Dynafleet online. Jeg ønsker at prosjektet skal være minst mulig til bry i en travel arbeidshverdag, og tar gjerne imot innspill for å sørge for dette.

Hva skjer med informasjonen om deg?
Alle personopplysninger vil bli behandlet konfidensielt. Jeg har tilgang til Dynafleet online mens prosjektet pågår, og overfører de aktuelle måledataene fra Dynafleet til egne filer hvor navn blir fjernet og dataene blir anonymisert. Veileder og bedriften får se anonymiserte data, slik som i den ferdige masteroppgaven. Navnelisten med telefonnummer og koblingsnøkkel for anonymiseringen oppbevares hver for seg på eksterne harddisker, og ikke sammen med anonymiserte data. Koblingsnøkkelen, navneliste og telefonnummer slettes når prosjektperioden er over. Prosjektperioden avsluttes ila juni.

Kun anonymiserte data lagres og brukes i masteroppgaven. Ingen personopplysninger publiseres. Bedriftens navn brukes i masteroppgaven. Dersom resultatene er interressante for videre publikasjon utover masteroppgaven, vil studenten kontakte deltakerne med egen forespørsel om dette.

Frivillig deltakelse
Det er frivillig å delta i studien, og du kan når som helst trekke ditt samtykke uten å oppgi noen grunn. Dersom du trekker deg, vil alle opplysninger om deg bli fjernet.

Dersom du har spørsmål til studien, ta kontakt med Aud K. Elnes, tlf. 930 64 192 eller epost audk@hotmail.com, eller veileder Elise F. Furrebøe tlf. 990 45 439 eller epost elise.furreboe@hioa.no.

For å ivareta god forskningsetikk er studien rutinemessig registrert hos Personvernombudet for forskning ved Norsk senter for forskningsdata (NSD).

Samtykke til deltakelse i studien
Jeg har mottatt informasjon om studien, og er villig til å delta.

(Signatur, dato, telefonnummer)

Appendix B. Average Idling With and Without Feedback.

		1	2	3	6	7	8	9	10
Phase B	FB	13,8 %	7,4 %	15,8 %	12,1 %	51,9 %	4,2 %	12,7 %	15,8 %
	No FB	16,9 %	9,7 %	19,5 %	8,1 %	56,8 %	6,4 %	9,5 %	19,4 %
	Difference:	-3,2 %	-2,3 %	-3,7 %	4,0 %	-4,9 %	-2,2 %	3,3 %	-3,6 %
Phase C	FB	14,2 %	9,3 %	13,6 %	12,2 %	28,8 %	3,8 %	13,3 %	11,9 %
	No FB	18,4 %	9,8 %	16,5 %	8,1 %	30,1 %	4,1 %	14,6 %	13,0 %
	Difference:	-4,2 %	-0,5 %	-2,9 %	4,1 %	-1,3 %	-0,2 %	-1,4 %	-1,1 %
Phase B+C	FB	13,9 %	7,8 %	15,0 %	12,1 %	42,4 %	4,1 %	12,8 %	14,6 %
	No FB	17,4 %	9,8 %	18,4 %	11,5 %	47,9 %	5,4 %	11,7 %	16,5 %
	Difference:	-3,4 %	-1,9 %	-3,3 %	0,6 %	-5,5 %	-1,3 %	1,2 %	-1,9 %

Appendix C. Number of Feedback Messages to Each Driver

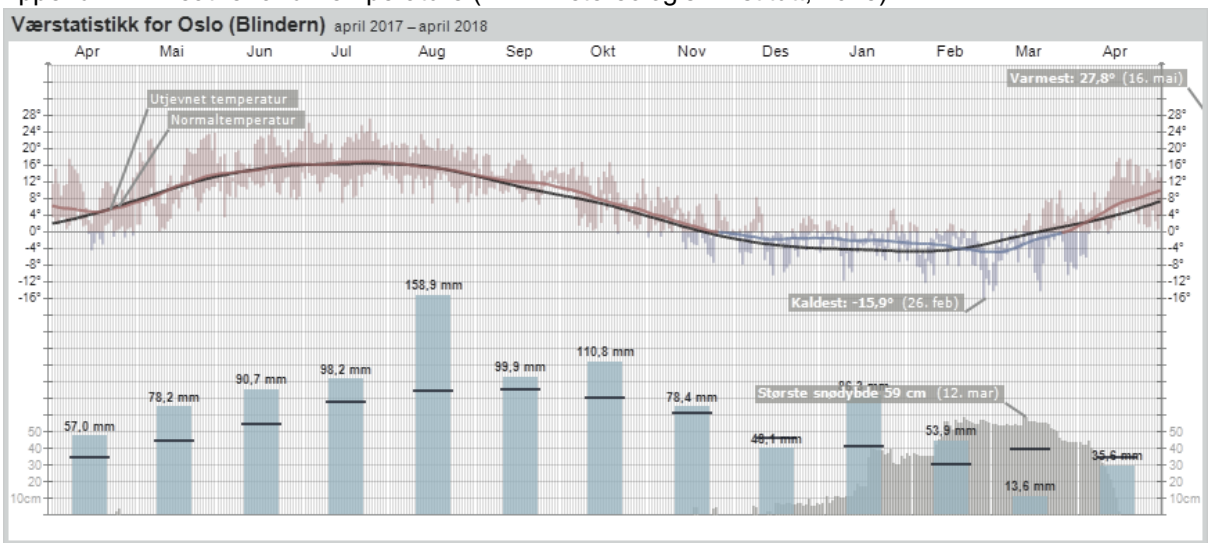
	1	2	3	6	7	8	9	10
Days feedback phase B	10	11	11	11	10	8	11	13
Days feedback phase B+C	17	14	17	13	18	12	15	19

Possible days of feedback was 19 in phase B and 11 in phase C, total of 30.

Appendix D. Correlation between Temperature and Idling

Correlation btw temperature and idling scores in data:									Withoout 4 and 5
	Driver 1	Driver 2	Driver 3	Driver 6	Driver 7	Driver 8	Driver 9	Driver 10	Group
Baseline	-0,11	0,19	-0,15	-0,29	0,05	-0,10	0,08	-0,07	-0,07
FB	0,04	0,32	0,41	0,41	-0,47	0,14	0,22	0,07	0,35
Info	-0,13	0,09	-0,48	0,28	-0,28	0,12	0,18	0,35	-0,07
FB + info	0,00	0,31	0,06	0,32	-0,68	-0,13	0,24	-0,21	-0,14
Dataset tota	-0,07	0,28	-0,08	0,05	-0,04	0,05	0,12	-0,07	0,05

Appendix E. Weather and Temperature (NRK Metereologisk Institutt, 2018)



Note. Average rain- or snowfall shown in the blue bars, snow measured each day shown in the grey area behind the blue bars in December to April.