



AARHUS UNIVERSITY



# Coversheet

---

**This is the accepted manuscript (post-print version) of the article.**

Contentwise, the post-print version is identical to the final published version, but there may be differences in typography and layout.

**How to cite this publication**

Please cite the final published version:

Holm, J. M. (2017). Double Standards? How Historical and Political Aspiration Levels Guide Managerial Performance Information Use. *Public Administration*, 95(4), 1026-1040.

<https://doi.org/10.1111/padm.12379>

## Publication metadata

<b>Title:</b>	Double Standards? How Historical and Political Aspiration Levels Guide Managerial Performance Information Use
<b>Author(s):</b>	Jakob Majlund Holm
<b>Journal:</b>	<i>Public Administration</i> , 95(4), 1026-1040
<b>DOI/Link:</b>	<a href="https://doi.org/10.1111/padm.12379">https://doi.org/10.1111/padm.12379</a>
<b>Document version:</b>	Accepted manuscript (post-print)

**General Rights**

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognize and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

**Double Standards? How Historical and Political  
Aspiration Levels Guide Managerial Performance  
Information Use.**

## **Abstract**

Performance evaluations are a vital part of how public managers react to performance information. Yet, we know little about the way evaluations unfold and the reactions they produce. This article examines the idea that historical and political aspiration levels guide evaluations by labeling performance results as either a failure or success. Thus, the same result could lead to different evaluations and reactions dependent on where on position of the aspiration level. The study also tests the idea that a negativity bias characterizes managers' reactions to political performance feedback. A survey-experiment among a sample of principals in Danish public schools is used to examine these hypotheses. The results show that aspiration levels change principals' interpretation and reaction to otherwise identical performance results, but primarily in situations with positive historical and political feedback. Conflicting feedback from two aspiration levels makes principals rely mostly on the signals from the historical aspiration level.

## **INTRODUCTION**

Evaluation of organizational performance is a key aspect of any managerial position. Managers need to evaluate performance to have a general awareness of organizational processes and because evaluations act as catalysts for a number of important decisions (Behn 2003; de Lancer Julnes 2006; Raudla 2012). The tools for performance evaluations have changed quite dramatically for public managers in recent years. Gone are the days where managers could only rely on gut feelings. Today, performance management systems support the process with performance information (Moynihan and Pandey 2010; Taylor 2011). While managers are supposed to integrate the information in their evaluations, we know little about how this process unfolds. This is unfortunate because the subjective assessments are the main driver of managerial action and organizational change (Meier et al. 2015).

The idea in this article is that aspiration levels guide how managers evaluate performance results and when they react to them. This claim originates from the behavioral model that was developed by March and Simon (1958) and Cyert and March (1963). In this model, managers are expected to form their evaluations by comparing absolute performance levels to an aspiration level, which indicates the least acceptable performance level (March and Simon 1958, p. 120). Thus, the main driver of managerial action is not absolute performance levels; rather it is discrepancies to the aspiration level in terms of negative and positive performance feedback. While this idea primarily has been tested in the private sector (see Shinkle 2012 for a review), recent research suggests that public managers also pay attention to aspiration levels (Salge 2011; Nielsen 2014; Rutherford and Meier 2015). However, the studies focus solely on the relationship between performance data and managerial decisions, thereby leaving a missing link on the information processing in-between. The aim of this study is to provide insights on this link by examining whether

aspiration levels (and performance feedback) influence managers' performance evaluations and their perceived need for managerial action.

The article contributes in three ways to our understanding of how managers react to performance information. As a first contribution, the study adapts the theoretical framework from the behavioral model to the public sector by considering the relevance of the political environment. While political goals in the form of performance targets have been shown to influence organizational attention and outcomes (Boyne and Chen 2006; Kelman and Friedman 2009), their role in managers' information processing remains uncertain. To understand this role, the study compares the relevance of political goals with a more generic aspiration, namely a historical aspiration level (i.e. the organization's past level of performance). Building on recent research, which shows that a negativity bias characterizes citizens' (Olsen 2017) and politicians' (Nielsen and Moynihan 2017) reactions to performance information, the expectation is that the consequential aspect of political feedback triggers asymmetric strong reactions to negative feedback. This hypothesis is tested with a survey-experiment among a group of principals in Danish public schools. The results confirm an asymmetric influence of negative and positive feedback. However, and contrary to the expectation, positive feedback has the most profound influence on performance evaluations made with both a historical and political aspiration level.

Second, the study utilizes the survey-experimental approach to separate the influence of aspiration levels from absolute performance, thereby providing causal evidence on the claim that aspiration levels change managers' interpretation and reaction to identical performance results. Furthermore, the analysis shows that principals' willingness to take action changes in accordance with their performance evaluation. Taken together, these findings add to the literature on performance information use (e.g. Kroll 2015; Andersen and

Moynihan 2016; Pandey 2016) by illustrating the cognitive processes underlying managerial reactions to the information.

The third contribution of the article concerns managerial reactions to conflicting signals on how well their organization is performing. While such ambiguity is often mentioned as a premise for performance information use (Moynihan 2008; Vakkuri 2010; Bækgaard and Serritzlew 2016), there is little theoretical and empirical basis for understanding the phenomenon. Multiple aspiration levels create an ambiguous situation when the comparisons indicate both a positive and a negative result. Managers could resolve this conflict either by collapsing the two levels into one (Olson et al. 1996) or by reflecting on each piece of feedback separately (Kahneman 1992). The results support the latter explanation. Principals' evaluations and reactions become less distinct in a situation with conflicting feedback. However, there is a skewed weighting of feedback from the two aspiration levels, the historical aspiration level being the dominate one.

## **THE COMPARATIVE NATURE OF EVALUATIONS**

A performance evaluation concerns the cognitive task of finding an answer to the fundamental managerial question; how well is my organization doing? To help answer this question, performance management systems deliver information on organizational outcomes to managers. This information reduces the task to interpretation and sense making of the results. However, an evaluation still entail a challenging aspect, as absolute numbers do not make much sense in themselves. For example, how is a principal to evaluate whether a school is well functioning when 65% of the parents are satisfied?

The key to this process is comparisons. This idea originates from research, which focus on either generic human traits (e.g. Kahneman and Tversky 1979; Mussweiler 2003; Köszegi and Rabin 2006) or more narrow on managerial information processing (e.g. Cyert

and March 1963; Greve 2008; Audia et al. 2015), and shows how judgements relies heavily on comparison to a standard, a norm or a particular piece of knowledge. For a manager who face a performance evaluation, the comparative aspect concerns the discrepancy between performance expectations and the actual outcome (Meier et al. 2015). In this way, expectations play a vital role in understanding how the evaluation unfolds.

One way managers could form these expectations is by consulting judgement-relevant knowledge (Nutley et al. 2012), for instance the amount of resources on the budget, characteristics of the clients, efficacy to affect the outcome, unforeseen events, and employee motivation. Recalling the hard-working employees probably increases expectations for a good result; while on the other hand, the thought of a challenging clientele would dampen the hope of success. Even though such pieces of knowledge support the formation of expectations, the evaluation remains rather complex as the ambiguous and none-numerical nature of the knowledge makes it difficult to establish a mental expression of a comparable number on the performance scale (Kahneman and Miller 1986).

As an alternative approach, the behavioral model suggests that managers use an aspiration level to form their expectations (Cyert and March 1963, p. 81). An aspiration level is a numerical representation of the least satisfactory performance level for the organization (Simon 1955, p. 103). Aspiration levels reduce the complexity of the evaluation because expectations reflect a single unified factor that is directly comparable to the absolute number, thereby enabling a simple categorization of the result as either a satisfactory or an unsatisfactory. In a situation where performance is above the aspiration level, the comparison creates positive feedback, which is a signal that the organization is performing better than expected (Simon 1979). Contrary, when performance is below the aspiration level, the comparison creates negative feedback (Cyert and March 1963, p. 39-40), indicating that the organization has failed the expectations.

While aspiration levels ease the cognitive effort in evaluations, the reliance on only one standard means that this number has a substantial influence on how the evaluation turns out. More specifically, if managers rely on aspiration levels to such degree that the one number triumphs other sources of knowledge and information, we should expect them to evaluate identical performance results differently dependent on whether an aspiration level is above or below the result. In the previous example, a principal might evaluate the satisfaction score of 65% as a fairly good result if the aspiration level was 60%. However, had the aspiration level been 85%, the negative discrepancy would lead the performance evaluation in a much more negative direction. Whether aspiration level cause such a dramatic shift in evaluations is tested in the first hypothesis.

*H<sub>1</sub>: Public managers evaluate performance result differently when presented with an aspiration level that is either above or below the absolute result.*

### **Aspiration levels in public organizations**

The concept of aspiration levels in an organizational setting was developed, and has primary been tested, in the private sector. This line of research has had focus on social (comparisons to similar organizations) and historical (comparisons to past performance) aspiration levels (see Shinkle 2012 for a review). These two types of aspiration levels have also characterized research in the public sector (Salge 2011; Nielsen 2014; Rutherford and Meier 2015; Olsen 2016). While there are generic traits between the private and public sector, which makes this application relevant (see Cyert and March 1963, p. 198), one important difference between the two sectors should be taken into consideration to fully understand how aspiration levels influence decision-making in the public sector. This difference is the political environment (Nutt 2005). The next section explains how this environment influences public managers' use

of aspiration levels and contrasts this influence to a more generic aspiration level, namely the historical aspiration level.

### **Historical and political aspiration levels**

A *historical aspiration level* is the organization's previous level of performance (Cyert and March 1963, p. 115). For example, take a hospital with an average waiting time pr. patient of 5 hours in 2015. When waiting time is measured again in 2016, the 5 hours is the historical baseline, which provides a standard of comparison for the performance evaluation. If waiting time had been measured to 3 hours pr. patient in 2016, the progression in performance would result in positive historical feedback (of 2 hours). A *political aspiration level* is a coercive and external set aspiration level in the form of a political goal, which define the level of performance that is expected by the political principals of the organization (Boyne and Chen 2006; Salge 2011). In contrast to historical aspiration levels where the comparison to the current performance level is eased by similarities in the metric scale, political aspiration levels vary in how precisely they express the aspiration level numerically (Chun and Rainey 2005). A goal that expresses the objective that a hospital's waiting time should reach a "satisfying" level requires a large degree of interpretation for a comparison to an absolute performance level. In contrast, if the goal includes a numerical target, for instance that the waiting time should be less than 2 hours, the process of making the comparison would be much easier.

Historical and political aspiration levels differ in one important way. While historical aspiration levels reflect actual organizational activities and outcomes, political aspiration levels are given exogenous and subjectively by political appointees. Accordingly, the two types of aspiration levels and their feedback are likely to reflect different things for managers. Historical aspiration levels are an aspiration to match past levels of performance

(Meier et al. 2015), which means that the feedback reflects negative and positive organizational development in relation to this aim. Political aspiration levels are based on the political ambitions for the organization in terms of how well it ought to be performing, and political feedback is a measure of goal attainment, which is often tied to rewards and sanctions as part of accountability systems (Hood and Dixon 2010). It should be noted, however, that even though the two types of aspiration levels include these distinct features, in some instances, they have an expression where the distinction is less clear. For example, in the situation where political aspirations levels are formed as a demand for performance improvement over time or when political interventions are based on decreasing organizational performance.

### **Asymmetric influence of negative and positive feedback**

The distinctive features of the two aspiration levels, and what their feedback mean to managers, are likely to influence how managers adjust their performance evaluations with the use of the aspiration levels. The magnitude of these adjustments is regulated by the manager's (unconscious) perception of how much negative feedback entails a failure, and how much positive feedback reflects organizational success.

A starting point for understanding this aspect of performance evaluations is to consider the possibility of asymmetric adjustments in the form a negativity bias. The negativity bias is a general tendency for individuals to react stronger to negative events, interactions, emotions and information in comparison to their positive counterparts (Rozin and Royzman 2001). This tendency have been found in emotions, perceptions, attention, learning patterns and memory (for a review see Baumeister et al. 2001), and recently, the negativity bias has also been shown to influence how politicians and citizens react to performance information. Low performance levels leads politicians to attribute a higher

degree responsibility for outcomes to managers (Nielsen and Moynihan 2016) and make citizens' performance evaluations become asymmetrical more negative (Olsen 2016). These strong responses to negative information are often explained as protection against future losses (Kahneman and Tversky 1979). For instance, a politician could lose an election by not reacting to bad performance results, and citizens could risk a loss in welfare if they choose a poorly performing service provider. An important feature of both these situations is that they entail a consequential aspect. In this way, the examples highlight an important aspect of the negativity bias, namely a context dependent component, which regulates the negative focus in accordance with the perceived consequences of a stimulus (Smith et al. 2006). If this tendency also characterizes managers' information processing, we should expect a particular strong reaction to negative feedback in instances where the failure entails a consequential aspect for either the organization or the manager.

Following this idea, the expectation is that a negativity bias will characterize comparisons made by a political aspiration level, while reactions to the historical aspiration level are not subject to the same degree of bias. The reason is that managers are expected to perceive a lack of goal achievement as especially problematic since the failure could have severe consequences for them and their organization. In fact, these consequences are often an institutionalized part of accountability systems, either through soft means (such as increased scrutiny) or more tangible sanctions in the form of financial punishment or firings (Jacob and Levitt 2003; Hood 2011; Soss et al. 2011; Martin et al. 2016). Even in systems where a lack of goal achievement does not automatically lead to consequences, the heavy focus on failure from politicians (Nielsen and Bækgaard 2015; Miller 2017) and citizens (Olsen 2017) could send the signal to managers that a performance gap in relation to political performance targets is highly problematic. This is not to say that a manager does not pay attention to negative historical feedback; rather the idea is that negative political feedback evokes a stronger

mental expression because the feedback is a focal point for (explicit or implicit) attention in performance management systems by informing on the progress towards the political goal (Moynihan 2008). In line with this argument, the expectation is that a negativity bias will characterize public managers' reactions to political feedback, while reaction to negative and positive historical feedback should be more balanced. These expectations are tested in H<sub>2.1</sub> and H<sub>2.2</sub>.

*H<sub>2.1</sub>: Negative political feedback make public managers adjust their performance evaluations downwards and positive political feedback leads to an upwards adjustment. The adjustments reflect a negativity bias.*

*H<sub>2.2</sub>: Negative historical feedback make public managers adjust their performance evaluations downwards and positive historical feedback leads to an upwards adjustment.*

### **Conflicting feedback**

So far, the argument has concerned three different scenarios, namely a situation with no aspiration level and the situations with either a historical or political aspiration level.

However, these situations do not fully capture the complexity in modern performance management systems. The data in these systems is often presented in such way that managers have both a historical and political aspiration level available in their information processing. An interesting aspect of this situation is evaluations where the signals are conflicting, as one type of feedback is positive, while the other is negative. How should we expect managers to react to such ambiguity?

A simply approach would be to combine the two aspiration levels to one an average level and make comparisons to this level (Feather 1966; Olson et al. 1996). The combined

level would move towards the absolute performance level when feedback is conflicting, which means that the comparison should leave the impression of a smaller performance gap (either negative or positive) than in a situation with only one aspiration levels. If the aspirations levels have the same numerical distance to the absolute number, no gap exists at all.

Another approach is separate processing of feedback from each aspiration level (Kahneman 1992; Ordonez et al. 2000). If managers have the cognitive capabilities to engage in this type of information processing, they register each piece of feedback and perceive the situation as both a success and failure at the same time (Kahneman 1992, p. 306). However, this does not necessarily mean that they place equally weight on negative and positive feedback in their evaluations. Two factors determine whether this is the case. First, distinctive features of the aspiration levels might lead one to dominate the other, no matter the content of the feedback (Chapman and Johnson 1999). Features that could make an aspiration level particular relevant is, for instance, familiarity, connectivity and internal locus (Yockey and Kruml 2009). Second, managers could use the aspiration levels strategically and form their preferences for the feedback in an endogenous process (Kahneman and Miller 1986). Strategic information processing could utilize both a positive and negative result. To place most emphasis on the positive result is consistent with self-enhancing tendencies that aim at maintaining a positive self-image as a competent manager (Jordan and Audia 2012). However, the focus could also shift towards the negative result as part of a manager's continuous striding for organizational survival through problem-solving (Simon 1952). While separate processing of the aspiration levels entails the idea of selective managerial reactions to aspiration levels and feedback, there are no compelling reasons to expect a scenario where managers completely ignore the less attractive aspiration level or feedback. Both historical and political aspiration levels fulfill criteria such as familiarity and

connectivity, and even in the case of strategic behavior, it would demand a high degree of ignorance to discard the undesirable result completely. In this way, both approaches expect managers to react to conflicting feedback by averaging across the aspiration levels, the difference being that the segregated approach entails the possibility of a weighted average, towards either a specific aspiration level or feedback. Because this average level is closer to the absolute performance result than in a situation with only one aspiration level, we should expect evaluations from conflicting feedback to have a less distinctive direction. Hypothesis 3 tests this argument.

*H<sub>3</sub>: When public managers' face conflicting feedback signals from two aspiration levels, their performance evaluations are based on (weighted) average between the two levels. Thus, the evaluations are less distinctive - compared to a situation with only one aspiration level.*

### **Managerial reactions to performance evaluations**

An important aspect of performance evaluations is how they guide managerial behavior. Building on the idea of aspiration levels, the overall expectation in the behavioral model is that negative feedback initiates a managerial reaction. This reaction involves a search for solutions, a choice of an appropriate solution, and implementation of this solution in order to correct the problems (Cyert and March 1963, p. 121). In contrast, positive feedback leaves a manager with a more room to maneuver and choose a course of action. As examples of actions, a manager could try to maintain the status quo (March and Shapira 1992) or reallocate resources from the successful performance dimension to other dimensions (Greve 2008). In the last category is an important distinction between using the resources to experiment and initiate innovative processes (Salge 2011) and using them for activities that does not directly benefit the overall purpose of the organization (i.e. slack) (Singh 1986).

Thus, while the expectation is that managers act on negative feedback, the process after a positive performance evaluation remains more uncertain.

## **DESIGN AND DATA**

### **An experimental approach**

The purpose of the study is to test how aspiration levels guide public managers' reactions to performance information. This question is challenging to study with the use of observational data because of endogeneity risks. Many different organizational and managerial variables could influence absolute performance, the presence and distance to an aspiration level and managers' reactions to the information. In addition, the relationship could suffer from reverse causality in the sense that information influences managers' reactions, but, over time, these reactions are also likely to influence performance (and thereby performance information). In order to eliminate these risks, the identification strategy for the study is a survey experiment conducted among a group of principals in Danish public schools.

### **Participants and experimental design**

The survey was sent online to the principals at 1225 Danish public schools on October 14, 2015. Three reminders were sent to the principals before the survey was closed on November 23, 2015. At the time, 158 principals had participated in the survey experiment, which equals a response rate of 13%. To improve the reliability of the questions, a pilot version of the survey was presented to four principals. An interview with these principals did not reveal any major flaws in the questions, as the principals only had minor suggestions for improvements.

The overall design of the experiment was to present principals with a case that described a school and its performance on four different areas. This multidimensionality was included to strengthen the ecological validity of the case as Principals in Denmark are

measured on a number of performance dimensions. The four areas were chosen because of their familiarity to the principals. GPA, truancy, and student well-being are all national measures of school performance, and parental satisfaction surveys are used in many municipalities as a supplement. Four versions of the case were created and principals were randomly assigned to one version. The cases were different in terms of the number and type of aspiration level(s) that were shown to principals - in addition to the school's absolute level of performance. Table 1 illustrates these differences.

Principals who were assigned to the *control group* were only shown the school's absolute level of performance for the four areas. The control group was designed to serve as a neutral baseline in order to test asymmetric reactions to negative versus positive feedback. Accordingly, the absolute performance level on the four performance areas was set to either a national average or an average from a representative municipality.

#### TABLE 1 ABOUT HERE

As shown in table 1, principals in treatment groups 1 and 2 were presented with one aspiration level for each performance area in addition to the absolute performance levels. The aspiration levels were set such that principals were shown two cases of negative feedback and two cases of positive feedback. It was randomized which areas were connected to the two types of feedback. For example, within a treatment group, one principal was shown negative feedback on the areas "GPA" and "Truancy", while another was shown a version with negative feedback on the areas "Truancy" and "Student well-being".

Principals in  $t_1$  (*the historical aspiration level*) were, in addition to the absolute level of performance, shown information about the school's performance on the four performance-areas in 2013. In two of the four areas, principals were shown a 25% decline in performance

from 2013 (performance areas one and four in table 1) and a 25% increase in performance in two areas (performance areas two and three in table 1). In the illustration in table 1, there is negative historical feedback on performance areas one and four and positive historical feedback on performance areas two and three.

This way of organizing the experimental condition was also used in  $t_2$  (*the political aspiration level*). Here, principals were told that the school was located in a municipality with a political goal for each of the four performance areas. The political goals expressed as a specific level of performance in the four areas. This level was set at +/- 25% of the absolute performance level. In table 1, the school has negative feedback on performance area two and three (performance 25% below the political goal) and positive feedback on performance area one and four (performance 25% above the political goal).

$t_3$  (*historical and political aspiration levels*) was designed to test principals' reactions to conflicting feedback. In each performance area, principals were shown two aspiration levels, a historical and a political, in addition to the absolute level of performance. When one aspiration level indicated positive feedback, the other aspiration level indicated negative feedback. For example, in the first performance area (GPA) in table 1, the historical aspiration level provides negative feedback (a performance decrease from 8.4 to 6.7) and the political aspiration level provides positive feedback (performance 25% above the political goal of 5). The treatment was designed so two performance areas showed this type of conflict, while the other two performance areas showed the reversed type of conflict (i.e. positive historical feedback and negative political feedback). As in  $t_1$  and  $t_2$ , the selection of a performance area to display a certain type of conflict was randomized.

### *Dependent variables*

After the principals were presented with one of the experimental groups, they were asked to assess each performance area. In order to ensure that principals' would react as realistically as possible, they were initially told to respond as they would in a similar situation with their own school. Specifically, principals were asked two questions:

#### *Performance evaluation*

How would you assess the school's performance on the four performance areas?

Response categories: 0-100

0 = Very poor

100 = Very well

#### *Need for managerial action*

To which degree do the results indicate a need for managerial action?

Response categories: 0-100

0 = No need

100 = A great need

### **Descriptive statistics and balance check**

Table 2 shows the number of principals in the four experimental groups along with a balance check for the four groups. As each principal was asked to react to four areas, there is a 1:4 ratio of principals to observations. This design creates a hierarchical structure with several observations nested within one principal. To account for this dependence between observations, all models are estimated with standard errors clustered on the principal.

TABLE 2 ABOUT HERE

As seen in table 2, there are fewer principals and observations in the control group compared to the treatment groups. This imbalance is a result of the two-step randomization procedure. Within a treatment group, the four different performance areas were randomly connected to negative and positive feedback, which technically could only be done by increasing the number of different conditions (within each treatment). An unfortunate result

of this procedure is that it lowered the probability of being assigned to the control group relative to the treatment groups. However, the imbalance is merely a matter of difference in the probability of being selected to the control groups relative to being selected to a treatment group (Green and Gerber 2012, p. 36). Principals were still assigned to a group through a random process, which means that the procedure did not create systematic differences across groups (as shown by the balance check in table 2).

Two points should be noted in relation to small sample in the control group. First, there is a statistical and substantial concern for the outcome means for this group. The statistical concern is that the combination of a small sample and a noisy outcome (i.e. with random measurement error) increases uncertainty in the estimates, and thereby the chance of obtaining an estimate that is larger than the true population value (Loke and Gelman 2017). Substantially, and in addition, for the experiment to be able to test managerial reactions to negative and positive feedback from a neutral point, the outcome-mean for the control group should be representative for a population of managers without aspiration levels.

Besides the fact that principals in the control group balance on covariates, two points speak in favor of valid outcome-means. Each principal makes four decisions on the two dependent variable, which means that the aggregated outcome-mean contains more information than the sole nine observations. In addition, the mean for control group is around 57 on the performance scale from 0-100 (for the performance evaluation), which is in line with the expectation as the absolute performance levels for the four performance areas were set to the national means.

Second, the small sample poses a challenge regarding inference and the ability to detect treatment differences between control and treatment groups. The classical way to estimate such differences depends on the sampling variability in the population (estimated with the use of standard errors) and a parametric distribution for the test statistic. In small

samples, these approximations may be problematic (Gerber and Green 2012, p. 115) and lead to overly high p-values and to narrow confidence intervals (Keele et al. 2012, p. 496). In order to check the validity of the classical approach to inference, the p-values in the main models (models 1.1 and 2.1 in table A1) have been robustness checked with the use of randomization inference. Randomization inference is a non-parametric approach where the p-values of the realized treatment-effect are obtained with the use of a sampling distribution created by estimates from all possible, or a sample of all possible, randomizations (Fisher 1935). The p-values from classic asymptotic inference and randomization inference are compared in table A3 in the appendix. While p-values from randomizations inference are slightly lower, the results of the two approaches are substantial identical.

Table 2 also shows a balance check across seven relevant covariates: two objective characteristics of the principal (experience and gender), two characteristics of the principals' leadership (involvement of teachers and autonomy), and three characteristics of the school (the share of students with a non-Danish ethnicity, the schools' geographical placement and the municipality's average income level). As mentioned, the balance check only shows minor differences between principals in the control group and the principals in the treatment groups. Furthermore, there are no significant imbalances between the treatment groups. Because of the minor imbalances, the main models without covariates (except for a control variable for the four performance areas) have been robustness checked with models that include covariates. Results from these models are presented in table A1 in the appendix. As shown by the table, there are no substantial differences between the two approaches.

Finally, the representativeness of the sample is worth mentioning. In relation to the principals' gender and experience, larger samples with response rates around 50% (Pedersen et al. 2011) show similar distributions to this sample. Here, the share of male principals is 62% (compared to 59% in the present sample) and the average years of experience is eight

(compared to 6.7 in the present sample). The average share of students with non-Danish ethnicity is 9.6% in the population, while it is 10.2% in the sample. In relation to municipality income, the average is 40.460 dkk. in the population versus an average of 39.490 dkk. in the sample. Thus, the principals in the sample do not substantially deviate from the population on these variables. Even if the principals in the sample deviates on some unmeasurable characteristics, the estimated treatment effects will be unbiased if the treatment effect is uncorrelated with the deviating variables (Druckman and Kam 2011).

Treatment effects are estimated by regressing the three treatment groups (split by the type of feedback) with the two outcomes. Estimated differences between treatment groups are presented in table A2.1 and A2.2 in the appendix. All results are robust to models where the groups are compared in pairs.

## **ANALYSIS**

Figure 1 illustrates the estimated differences between the control group and the treatment groups for the two outcomes. The first notable result in the figure is that unambiguous negative and positive feedback changes principals' performance evaluations. This result is in line with the expectation in  $H_1$ , namely that aspiration level changes public managers' evaluations of identical performance level.

### **FIGURE 1 ABOUT HERE**

Figure 1 also shows that principals adjust their performance evaluations in an asymmetric manner. However, the asymmetries are not as expected in  $H_{2.1}$  and  $H_{2.2}$ . Contrary to the expectation in  $H_{2.1}$ , there is no negativity bias in relation to political feedback. In fact, negative political feedback has no influence on principals' performance evaluations at all.

This result is in sharp contrast to the substantial influence of positive political feedback. Positive political feedback increases an evaluation in a positive direction by as much as 28-scale points. For historical feedback, the asymmetries between negative and positive feedback are more balanced, as expected in H<sub>2.2</sub>. Negative historical feedback changes performance evaluations in negative direction by 13 scale points and positive historical feedback influence evaluations positively by 20 scale points. Thus, there is also a positive asymmetry for this type of feedback, however, to a lesser degree than for the political aspiration level.

Substantially, the results point to two interesting tendencies in the way principals rely on aspiration levels. First, reactions to aspiration level can be highly selective. This is especially illustrated by the fact that performance below a political goal is not perceived as particularly problematic; while goal achievements is perceived as a great success. In this way, principals use a double standard in their reliance on political feedback. Second, it is worth reminding that negative historical feedback signalizes a performance decrease of 25%. It is interesting that such a substantial performance decline influences principals' reactions only in to a minor degree. This point is particular evident when examining the right side of figure 1, which show principals' perceptions of whether a performance area requires managerial action. By comparing these results to the performance evaluations on the left side, it is evident that even though principals to some degree acknowledge the performance decline in their evaluation, they do not believe the result increases the need for managerial action. Actually, the principals' logic in their reliance of aspiration levels is the exact opposite; for both positive historical and political feedback, they use their positive performance evaluation to reduce attention on the successful performance area.

The bottom of figure 1 shows how principals' performance evaluations change when they experience conflicting feedback. The findings from this part of the analysis partly

confirms  $H_3$  and the expectation that conflicting feedback makes evaluations less distinct than in a situation with only one aspiration level. To nuance the first-hand impression in figure 1, additional analysis are provided in table A2.1 and A2.2 in the appendix. These estimates indicate that even though evaluations become less distinctive (and resemble those made in a situation without an aspiration level), the reduction is skewed towards the signal from the historical aspiration level. Evaluations with historical feedback change marginally, going from a situation with lone feedback to conflicting feedback. The numerical changes are eight and six scale-points for positive and negative feedback. For the political aspiration level, the story is the opposite. Principals reduce their performance evaluations by as much as 33 scale points ( $p$ -value  $< 0.001$ ) when the positive signal from the political feedback is challenged by a negative signal from the historical feedback. The same tendency is evident for negative political feedback where principals' performance evaluations become 15 percentage points ( $p$ -value  $< 0.001$ ) more positive when they receive positive historical feedback in addition to the feedback from the political target.

## **DISCUSSION AND CONCLUSION**

This article answers two important questions about how historical and political aspiration levels guide public managers' reactions to performance information. First, the analysis show that a sample of public school principals from Denmark interpret performance results differently when an aspiration level is present, indicating either negative or positive performance feedback. The principals change their evaluations asymmetrically in relation to feedback from both a historical and political aspiration level. Positive feedback influences evaluations the most. This type of feedback make principals' evaluations of performance more positive and lowers their willingness to take action on the successful performance dimension. Second, the study provides evidence on how managers react in a situation with

ambiguous signals from multiple aspiration levels. In this situation, principals mostly rely on signals from the historical aspiration level.

These conclusions were established in a survey-experiment, which has the advantage that effects can be given a causal interpretation. Still, two limitations to the study should be noted. First, the survey-experiment can only replicate the context where principals process and react to performance information to a certain degree. Despite the ecological validity of the treatment, principals act in a more complex reality when they use performance information as part of their daily management. On the other hand, the elimination of contextual factors is also an advantage of the study because the results provide an account of how much managers rely on performance information without any environmental input and incentives.

Second, to answer a survey is an exercise with low stakes for the principals. The lack of a finding in relation to negative political feedback does not necessarily mean that managers will not react to this type of feedback when stakes are much higher; for instance in a meeting with a political principal. Following this line of reasoning, we should not completely discard negativity bias as a factor, which could influence managers' reactions to performance information. Instead, future research should try to detangle how different political contexts, incentives and performance management systems influence whether managers react to negative information signals.

In a broader perspective, the results provide an answer to the important question of why managers seldom react to performance information (Moynihan and Pandey 2010), namely because their performance evaluations do only not reflect the negative content in the information. In this way, the study questions the idea of a problem-solving manager who uses performance information to identify and correct organizational problems as a way to improve performance (Cyert and March 1963). On the other hand, if managers mostly pay attention to

positive performance results, it naturally raises the question of what this focus entails. An important aspect for future research is whether managers actually utilize positive feedback to learn of functional organizational practices (Argote and Miron-Spektor 2011) or if the focus reflects less purposeful processes such as self-enhancement (Jordan and Audia 2010) and initiation of slack-activities (Singh 1986).

## **Acknowledgements**

I am grateful to Simon Calmar Andersen and Mads Leth Jakobsen, three anonymous reviewers, and the participants at the panel for Behavioral and Experimental Public Administration at the IRSPM 2016 Hong Kong Conference for their helpful comments and suggestions on earlier drafts of this article. Correspondence to the author at [jholm@ps.au.dk](mailto:jholm@ps.au.dk).

## REFERENCES

- Andersen, S. C. and D. P. Moynihan. 2016. 'How Leaders Respond to Diversity: The Moderating Role of Organizational Culture on Performance Information Use', *Journal of Public Administration Research and Theory*, 26, 3, 448–460.
- Argote, L. and E. Miron-Spektor. 2011. 'Organizational Learning: From Experience to Knowledge', *Organization Science*, 22, 5, 1123-1137.
- Audia, P G., S. Brion and H. R. Greve. 2015. 'SelfAssessment, Self-Enhancement, and the Choice of Comparison Organizations for Evaluating Organizational Performance', *Cognition and Strategy*, 32, 89-118.
- Baekgaard, M. and S. Serritzlew. 2016. 'Interpreting Performance Information: Motivated Reasoning or Unbiased Comprehension', *Public Administration Review*, 76, 1, 73–82.
- Baumeister, R. F., E. Bratslavsky, C. Finkenauer and K. D. Vohs. 2001. 'Bad is Stronger Than Good', *Review of General Psychology*, 5, 4, 323-370.
- Behn, R. D. 2003. 'Why Measure Performance? Different Purposes Require Different Measures', *Public Administration Review*, 63, 5, 586–606.

Boyne, G. A. and A. A. Chen. 2006. 'Performance Targets and Public Service Improvement', *Journal of Public Administration Research and Theory*, 17, 3, 455-477.

Chapman, G. B. and E. J. Johnson. 1999. 'Anchoring, Activation, and the Construction of Values', *Organizational Behavior and Human Decision Processes*, 79, 2, 115–153.

Chun, Y. H. and H. G. Rainey. 2005. 'Goal Ambiguity in U.S. Federal Agencies', *Journal of Public Administration Research and Theory*, 15, 1, 1-30.

Cyert, R. M. and J. G. March. 1963. *A Behavioral Theory of the Firm*. Englewood Cliffs, N.J: Prentice-Hall.

de Lancer Julnes, P. 2006. 'Performance Measurement – An Effective Tool for Government Accountability? The Debate Goes On', *Evaluation*, 12, 2, 219–235.

Druckman, J. N. and C. D. Kam. 2011. 'Students as Experimental Participants: A Defense of the "Narrow Data Base"' in J. N. Druckman, D. P. Green, J. H. Kuklinski and A. Lupia (eds.), *Cambridge Handbook of Experimental Political Science*. New York: Cambridge University Press, 41–57.

Feather, N. T. 1966. 'Effects of Prior success and Failure on Expectations of Success and Subsequent Performance', *Journal of Personality and Social Psychology*, 3, 3, 287-298.

Fisher, R. A. 1935. *The Design of Experiments*. London: Oliver and Boyd.

Green, D. P. and A. S. Gerber 2012. *Field Experiments: Design, Analysis, and Interpretation*. WW Norton.

Greve, H. R. 2008. 'A Behavioral Theory of Firm Growth: Sequential Attention to Size and Performance Goals', *Academy of Management Journal*, 51, 3, 476-494.

Hood, C. and R. Dixon. 2010. 'The Political Payoff From Performance Target System: No-Brainer or No-Gainer?', *Journal of Public Administration Research and Theory*, 20, 2, 281–291.

Hood, Christopher. 2011. *The blame game: Spin, bureaucracy, and self-preservation in government*. Princeton, NJ: Princeton University Press.

Jacob, B. A. and S. D. Levitt. 2003. 'Rotten Apples: An Investigation of the Prevalence and Predictors of Teacher Cheating', *Quarterly Journal of Economics*, 118, 3, 843–877.

Jordan, A. H. and P. G. Audia. 2012. 'Self-enhancement and Learning From Performance Feedback', *Academy of Management Review*, 37, 2, 211–231.

Kahneman, D. 1992. 'Reference Points, Anchors, Norms, and Mixed Feelings', *Organizational Behavior and Human Decision Processes*, 51, 2, 296-312.

Kahneman, D. and A. Tversky. 1979. 'Prospect Theory: An Analysis of Decision Under Risk', *Econometrica*, 47, 2, 263-292.

Kahneman, D. and D. T. Miller. 1986. 'Norm Theory: Comparing Reality to its Alternatives', *Psychological Review*, 93, 2, 136-153.

Keele, L. J., C. McConaughy and I. K. White. 2012. 'Strengthening the Experimenter's Toolbox: Statistical Estimation of Internal Validity', *American Journal of Political Science*, 56, 2, 484-99.

Kelman, S. and J. N. Friedman. 2009. 'Performance Improvement and Performance Dysfunction: An Empirical Examination of Distortionary Impacts of the Emergency Room Wait-time Target in the English National Health Service', *Journal of Public Administration Research and Theory*, 19, 4, 917-46.

Koszegi B. and M. Rabin. 2006. 'A Model of Reference-Dependent Preferences', *Quarterly Journal of Economics*, 12, 4, 1133-65.

Kroll, A. 2015. 'Drivers of Performance Information Use: Systematic Literature Review and Directions for Future Research', *Public Performance and Management Review*, 38, 3, 459-486.

Loken, E. and A. Gelman. 2017. 'Measurement Error and the Replication Crisis', *Science*, 355, 584-585.

March, J. G. and H. A. Simon. 1958. *Organizations*. Cambridge, MA: Blackwell.

March, J. G. and Z. Shapira. 1992. 'Variable Risk Preferences and the Focus of Attention', *Psychological Review*, 99, 1, 172-183.

Martin, S., S. Nutley, J. Downe and C. Grace. 2016. 'Analysing Performance Assessment in Public Services: How Useful is the Concept of a Performance Regime?', *Public Administration*, 94, 1, 129-45.

Meier, K. J., N. Favero and L. Zhu. 2015. 'Performance Gaps and Managerial Decisions: A Bayesian Theory of Why Managers Seek Information and Take Risks', *Journal of Public Administration Research and Theory*, 25, 4, 1221-1246.

Miller, S. M. 2017. 'How Electoral Incentives Shape Performance Measurement Efforts', *Public Administration*, 95, 3, 623-637.

Moynihan, D P. 2008. *Dynamics of performance management: Constructing information and reform*. Georgetown University Press.

Moynihan, D. P. and S. K. Pandey. 2010. 'The big question for performance management: Why do managers use performance information?', *Journal of Public Administration Research and Theory*, 20,4, 849-66.

Mussweiler, T. 2003. 'Everything is Relative: Comparison Processes in Social Judgment', *European Journal of Social Psychology*, 33, 6, 719 -733

Nielsen, P. A. 2014. 'Learning From Performance Feedback: Performance Information, Aspiration Levels, and Managerial Priorities', *Public Administration*, 92, 1, 142–60.

Nielsen, P. A. and D. P. Moynihan. 2017. 'How do Politicians Attribute Bureaucratic Responsibility for Performance? Negativity Bias and Interest Group Advocacy', *Journal of Public Administration Research and Theory*, 27, 2, 269-283.

Nielsen, P. A. and M. Bækgaard. 2015. 'Performance Information, Blame Avoidance, and Politicians' Attitudes to Spending and Reform: Evidence from an Experiment', *Journal of Public Administration Research and Theory*, 25, 2, 545–69.

Nutley, S., R. Levitt, W. Solesbury and S. Martin. 2012. 'Scrutinizing Performance: How Assessors Reach Judgements about Public Services', *Public Administration*, 90, 4, 869–85.

Nutt, P.C. 2005. 'Comparing Public and Private Sector Decision Making Practices', *Journal of Public Administration Research and Theory*, 16, 2, 289-318.

Olsen, A. L. 2017. 'Compared to What? Experimental Evidence on Social and Historical Reference Points in Performance Evaluation', *Journal of Public Administration Research and Theory*, 27, 4, 562-580.

Olson, J. M., N. J. Roese and M. P Zanna. 1996. 'Expectancies' in E. T. Higgins and A. W. Kruglanski (eds.) *Socialpsychology: Handbook of Basic Principles*. New York: Guilford Press, 211-238.

Ordóñez, L.D., T. Connolly and R. Coughlan. 2000. 'Multiple Reference points in Satisfaction and Fairness Assessment', *Journal of Behavioral Decision Making*, 13, 3, 329-344.

Pandey, S. P. 2016. 'Performance Information Use: Making Progress But a Long Way to Go', *Public Performance & Management Review*, 39, 1, 1-6.

Pedersen, M. J., A. Rosendahl, S. C. Winter, A. P. Langhede and M. Lynggard. 2011. *Ledelse af Folkeskoler: Vilkår og Former for Skoleledelse*. SFI - National Research Center for Welfare.

Raudla, R. 2012. 'The Use of Performance Information in Budgetary Decision-Making by Legislators: Is Estonia any Different?', *Public Administration*, 90, 4, 1000–1015.

Rozin, P. and E. B. Royzman. 2001. 'Negativity Bias, Negativity Dominance, and Contagion', *Personality and Social Psychology Review*, 5, 4, 296-320.

Rutherford, A. and K. J. Meier. 2015. 'Managerial Goals in a Performance-Driven System: Theory and Empirical Tests in Higher Education', *Public Administration*, 93, 1, 17–33.

Salge, T. O. 2011. 'A Behavioral Model of Innovative Search: Evidence from Public Hospital Services', *Journal of Public Administration Research and Theory*, 21, 1, 181–210.

Shinkle, G. A. 2012. 'Organizational Aspirations, Reference Points, and Goals: Building on the Past and Aiming for the Future', *Journal of Management*, 38, 1, 415-455.

Simon, H. A. 1955. 'A Behavioral Model of Rational Choice', *Quarterly Journal of Economics*, 69, 1, 99-108.

Simon, H. A. 1979. 'Rational Decision Making in Business Organizations', *American Economic Review*, 69, 4, 493–513.

Singh, J.V. 1986. 'Performance, Slack, and Risk Taking in Organizational Decision-making', *Academy of Management Journal*, 29, 3, 562–585.

Smith, N. K., J. T. Larsen, T. L. Chartrand, J. T. Cacioppo, H. A. Katafiasz and K. E. Moran. 2006. 'Being Bad isn't Always Good: Affective Context Moderates the Attention Counter-Regulation Toward Negative information', *Journal of Personality and Social Psychology*, 90, 2, 210- 220.

Soss, J., R. Fording and S. F. Schram. 2011. 'The Organization of Discipline: From Performance Management to Perversity and Punishment', *Journal of Public Administration Research and Theory* 21, 2, 203–32.

Taylor, J. 2011. 'Strengthening the Link Between Performance Measurement and Decision Making', *Public Administration*, 87, 4, 853–71.

Vakkuri, J. 2010. 'Struggling with Ambiguity: Public Managers as Users of NPM-Oriented Management Instruments', *Public Administration*, 88, 4, 999 – 1024.

Yockey, M. D. and S.M. Kruml. 2009. 'Everything is Relative, but Relative to What? Defining and Identifying Reference Points', *Journal of Business and Management*, 15, 1, 95-109.

## TABLES AND FIGURES

TABLE 1: *Illustration of the four experimental conditions.*

	<b>Result in 2013</b>	<b>Result</b>	<b>Political goal</b>
Grade point average	8.4	6.7	5.0
Share of satisfied parents	56 %	75 %	94 %
Student truancy	7.1 %	5.7 %	4.3 %
Share of satisfied students	100 %	80 %	60 %
<i>c</i> : No aspiration level		X	
<i>t</i> <sub>1</sub> : Historical aspiration level	X	X	
<i>t</i> <sub>2</sub> : Political aspiration level		X	X
<i>t</i> <sub>3</sub> : Historical and political aspiration level	X	X	X

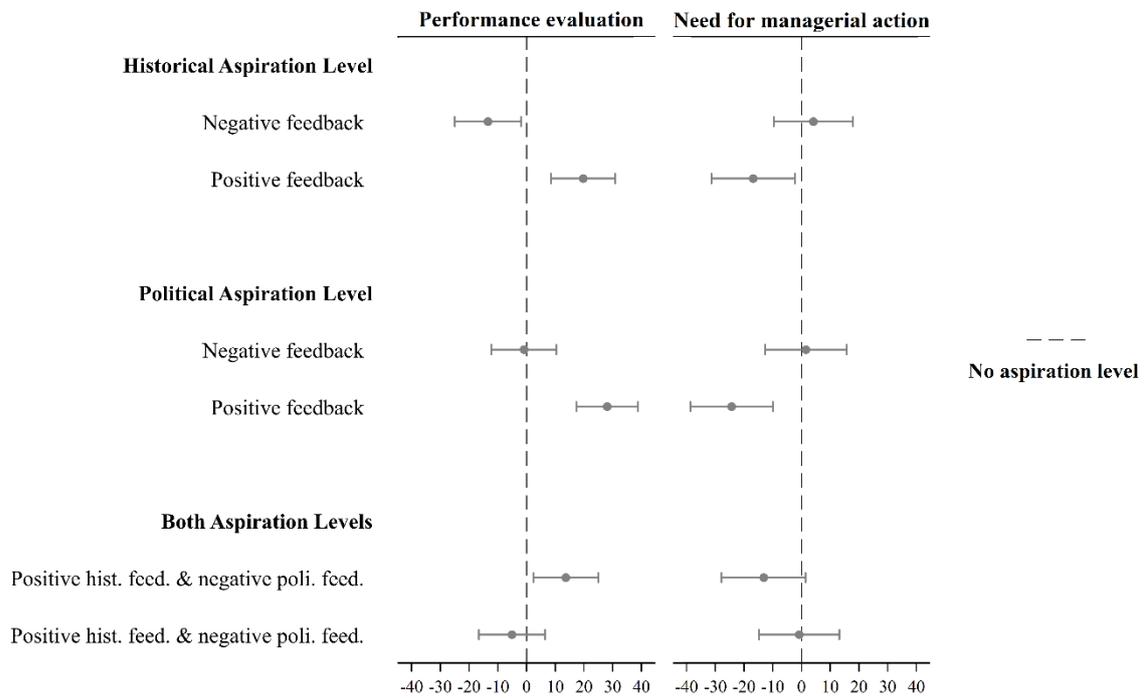
Note: The table illustrates one way negative and positive feedback was randomized in the treatment groups.

TABLE 2: *Descriptive statistics and balance check for the four experimental-groups.*

	Control	Treatment <sub>1</sub>	Treatment <sub>2</sub>	Treatment <sub>3</sub>
Principals	9	52	52	45
Observations	36	208	208	179
The principal's experience:	6.5	7.4	7.1	6.4
Average number of years at the school				
Schools with a male principal	78%	65%	63%	62%
Average percentage of students with non-Danish ethnicity	10%	11%	13%	8%
Schools geographically placed on the mainland	67%	60%	65%	73%
Schools with decreased autonomy during the last two years <sup>1</sup>	33%	34%	35%	36%
Schools where teachers have a saying in the teaching/preparation ratio <sup>1</sup>	56%	52%	49%	60%
Average income in the municipality (in dkk.)	40949 <sup>†</sup>	38506	39357	39374

Note: <sup>†</sup>  $p < 0.10$  (two-sided t-test) between control group and treatment group<sub>1</sub>. <sup>1</sup> Based on survey questions. All combination of groups have been tested using a two-sample t-test.

FIGURE 1: *How aspiration levels and performance feedback influence principals' reactions to performance information. OLS regression coefficients with 90% confidence intervals.*



Note: The figure is based on models 1.1 and 2.1 from table A1 in the appendix. Standard errors are clustered on the principal.

## APPENDIX

TABLE A1: *How aspiration levels/performance feedback influence principals' performance evaluations and perceived need for managerial action. OLS with clustered standard errors.*

Dependent variable	Performance evaluation		Need for managerial action	
	1.1 <sup>1</sup>	1.2 <sup>2</sup>	2.1 <sup>1</sup>	2.2 <sup>2</sup>
No aspiration level	Reference	Reference	Reference	Reference
Negative historical feedback	-13.4 <sup>†</sup> (-1.93)	-12.1 <sup>†</sup> (-1.91)	4.14 (0.50)	2.68 (0.44)
Positive historical feedback	19.7** (2.91)	21.1** (3.39)	-16.8 <sup>†</sup> (-1.91)	-18.1** (-2.68)
Negative political feedback	-0.88 (-0.13)	1.62 (0.26)	1.55 (0.18)	-1.01 (-0.16)
Positive political feedback	28.1*** (4.33)	30.6*** (5.18)	-24.3** (-2.80)	-26.9*** (-4.13)
Positive hist. feed. & negative poli. feed.	13.7* (2.01)	17.0** (2.77)	-13.2 (-1.48)	-15.9* (-2.43)
Positive poli. feed. & negative hist. feed.	-5.08 (-0.73)	-1.78 (-0.28)	-0.77 (-0.09)	-3.44 (-0.56)
Constant	56.8*** (8.68)	67.5*** (5.64)	76.0*** (9.26)	72.7*** (4.61)
<i>N</i> (principals)	158	158	158	158
<i>N</i> (observations)	631	631	628	628
Adjusted R <sup>2</sup>	0.33	0.36	0.16	0.18
Controls	No	Yes	No	Yes

Note: <sup>†</sup> = p<0.10; \* = p<0.05; \*\* = p<0.01; \*\*\* = p<0.001 (two-sided test). t-statistics in parenthesis. <sup>1</sup> Includes control for performance area. <sup>2</sup> Respondents with valid responses on treatment and outcome but missing values on covariates are included through the missing-indicator method.

TABLE A2.1: *How conflicting signals changes principals' performance evaluations.*  
*Predicted values and estimated difference.*

	<b>Lone</b>	<b>+ Conflicting signal</b>	<b>Difference</b>
Negative historical feedback	42.84	51.20	8.36* (2.06)
Positive historical feedback	75.98	69.97	-6.01 <sup>†</sup> (1.76)
Negative political feedback	55.40	69.97	14.57*** (4.10)
Positive political feedback	84.36	51.20	-33.16*** (-10.52)

Note: <sup>†</sup> = p<0.10; \* = p<0.05; \*\*\* = p<0.001 (two-sided test). t-statistics in parenthesis. The table is based on linear combinations of parameters from model 1.1 in table A1.

TABLE A2.2: *How conflicting signals changes principals' perceived need for managerial action. Predicted values and estimated difference.*

	<b>Lone</b>	<b>+ Conflicting signal</b>	<b>Difference</b>
Negative historical feedback	79.00	74.10	-4.91 (-1.57)
Positive historical feedback	58.06	61.71	3.65 (0.72)
Negative political feedback	76.41	61.71	-14.70** (-3.18)
Positive political feedback	50.55	74.10	23.55*** (5.89)

Note: \*\*=p<0.01; \*\*\*=p<0.001 (two-sided test). t-statistics in parenthesis. The table is based on linear combinations of parameters from model 2.1 in table A1.

TABLE A3: *Estimation of treatment effects using asymptotic and randomization inference. P-values.*

<b>Dependent variable</b>	<b>Performance evaluation</b>		<b>Need for managerial action</b>	
	<b>Asymp-totic</b>	<b>Randomi-zation</b>	<b>Asymp-totic</b>	<b>Randomi-zation</b>
Negative historical feedback	0.056	0.041	0.619	0.574
Positive historical feedback	0.004	0.002	0.058	0.021
Negative political feedback	0.898	0.903	0.857	0.821
Positive political feedback	0.000	0.000	0.006	0.001
Positive hist. feed. / negative poli. feed.	0.046	0.047	0.140	0.095
Negative his. feed. / positive poli. feed.	0.469	0.525	0.927	0.914

Note: Randomization inference is simulated with a sample of 1000 randomizations. Comparisons are based on models 1.1 and 2.1 in table A1.