

Nonresponse bias in public leadership research: An empirical assessment

Dominik Vogel^{a*} and Christian B. Jacobsen^b

^aDepartment of Socioeconomics, University of Hamburg, Germany; ^bDepartment of Political Science, Aarhus University, Denmark

* Dominik Vogel, University of Hamburg, Von-Melle-Park 9, 20146 Hamburg, dominik.vogel-2@uni-hamburg.de

This is an Accepted Manuscript of an article published by Taylor & Francis in the *International Public Management Journal* on 2021-04-09, available online:

<https://www.tandfonline.com/10.1080/10967494.2021.1906803>.

NONRESPONSE BIAS IN PUBLIC LEADERSHIP RESEARCH: AN EMPIRICAL ASSESSMENT

Self-reported measures of leadership are widely used in public management research, but nonresponse bias poses a threat to the validity of these data. Although this measurement problem is acknowledged, it has received limited empirical attention because nonresponse bias is inherently challenging to study. To address this issue, we examine nonresponse bias among public managers by analyzing multilevel surveys of managers and employees in which we can compare employee ratings of leadership for both responding and nonresponding managers. Using 16,531 employee responses spread over six datasets from three countries, we find only limited evidence of nonresponse bias in managers' self-reported leadership. Additional Bayesian analyses indicate that—overall—the data are indicative of the absence of substantial nonresponse bias. However, the results vary between datasets and call for more research on nonresponse bias in leadership research.

Keywords: nonresponse, leadership, nonresponse bias

INTRODUCTION

Information about public leadership is primarily collected through survey methods (Meier and O'Toole 2011; Knies, Jacobsen, and Tummers 2016). Surveys are useful for several reasons, including the ability to obtain information about variables that are difficult to observe directly, the consistency of measurement within and across studies, and the efficiency of gathering data on a large number of leaders or followers. However, in survey data, nonresponse is always an issue, and we know almost nothing about the representativeness of the collected data in leadership studies. Nonresponse is an important issue because selection bias in survey-based leadership studies can threaten the inferences made from such studies (Angrist and Pischke 2010; Kreuter, Muller, and Trappmann 2010). The literature on nonresponse in surveys has shown that factors such as interest in the survey topic and motivation can not only influence response rates (Armstrong and Overton 1977) but also lead to nonresponse bias. Thus, leaders who are motivated and interested in leadership may be more prone to respond to surveys, resulting in bias in survey-based leadership studies. However, we have scarce empirical knowledge on the existence and scope of such nonresponse bias in leadership studies – and public management in general.

The literature on leadership in the public sector has focused on different types of leadership, such as transformational, transactional, and servant leadership, which have been measured through surveys administered to leaders and their subordinates (Jacobsen and Andersen 2015; Oberfield 2014; Schwarz et al. 2016). Recent studies have also suggested specific aspects of public leadership and developed survey instruments to

obtain information about these leadership behaviors (Tummers and Knies 2016; Vogel, Reuber, and Vogel 2020). The aim of our study is not so much to focus on specific aspects of leadership but rather to investigate how self-reported measures of leadership, in general, can be affected by nonresponse bias. We expect that various aspects of leadership are related to survey response patterns because they reflect interest, motivation, and capability to exert leadership and, therefore, also to reply to surveys about leadership.

Nonresponse bias has been investigated in the survey methodology literature (Groves 2006; Groves et al. 2002) but not specifically concerning leadership. Therefore, this study's important contribution is to address the role of nonresponse and provide empirical insight into the magnitude and importance of nonresponse bias in leadership research. Furthermore, we suggest a method for studying nonresponse bias in leadership research that addresses differences between responding and nonresponding managers by looking at differences in their followers' survey responses.

To do so, we collected six multilevel datasets from three different countries that all consist of survey responses from both managers and their employees. These datasets from Denmark, Germany, and the Netherlands were collected from different areas of the public sector and cover both classical bureaucratic settings in cities and state agencies as well as frontline service organizations, such as schools, daycare centers, and welfare offices. In total, we analyze responses from 16,531 followers and 1,164 leaders. To preview our results, we find only limited indications of nonresponse bias. In some cases, responding managers are perceived by their followers as exerting more leadership than

nonresponding leaders. And we also find some differences related to particular leadership strategies and areas. But because these findings come from very different samples, settings and measures, we remain cautious in drawing firm conclusions with regard to the extent of nonresponse bias generally in public leadership research.

THE ISSUE OF NONRESPONSE BIAS IN SURVEYS

Nonresponse is a potential threat in survey research, where data is usually generated based on voluntary respondent participation (Edwards et al. 2009). Nonresponse is considered important in organizational research, and studies generally have experienced declining response rates (see, e.g., Pedersen and Nielsen 2016; Bickart and Schmittlein 1999; de Leeuw and de Heer 2002), which is also to some extent explained by the switch from mail and telephone surveys to online surveys (Manfreda et al. 2008). Still, the attainment of high response rates (low levels of nonresponse) is considered an essential aspect of survey quality (Pedersen and Nielsen 2016) because, *ceteris paribus*, low nonresponse increases the generalizability of the results. However, nonresponse is not necessarily a problem if respondents are similar to nonrespondents, in which case it becomes only an issue of statistical power (sample size). However, if nonrespondents and respondents differ on aspects related to the study's questions, nonresponse can produce bias in the study's findings.

Nonresponse bias has been defined as the magnitude of the difference in the answers by responders and nonresponders multiplied by the proportion of nonrespondents (Lambert and Harrington 1990, 5):

$$\text{Nonresponse bias} = \text{Nonresponse rate} \times (\text{Respondent estimate} - \text{Nonrespondent estimate})$$

According to this view, the level of nonresponse in a survey is important because it reduces nonresponse bias, but having a representative set of respondents is an even more important driver for avoiding nonresponse bias. An obvious challenge to assessing the level of nonresponse bias is that only the response rate (the level of nonresponse) and the respondent estimates are observable, whereas the nonrespondent estimate is almost always unobservable. Nonresponse bias is problematic because it is a special case of selection bias, which challenges the potential for drawing inferences from empirical studies (Angrist and Pischke 2014). Thus, if a study based on a given sample is hampered by nonresponse bias, we cannot trust that the results are representative of the whole population. Hence, nonresponse can cause several problems, such as reduced generalizability of the research findings, bias in sample estimates, and loss of information (e.g. Adigüzel and Wedel 2008).

What can cause nonrespondents to differ from respondents? Early studies on nonresponse bias have shown that factors such as interest and motivation increase response rates (Armstrong and Overton 1977) but that these factors are at the same time drivers of nonresponse bias. More specifically, the leverage-salience theory revolves

around the argument that the propensity to participate in surveys varies between individual respondents and depends on the presentational content of the survey request (Groves, Singer, and Corning 2000). Various aspects of the presentational content, such as the topic, the layout, or the length of the survey, will appeal differently to potential respondents, affecting the response pattern. The leverage-salience theory is a general framework that can provide a better understanding of how leverage matters for response patterns. For example, Lahaut et al. (2002) have shown how both non-drinkers and heavy alcohol users are less likely to participate in surveys about alcohol consumption, biasing results. In a meta-analysis on nonresponse rates and nonresponse bias, Groves and Peytcheva (2008) conclude that design features and characteristics of the sample can be predictors of nonresponse bias. Respondents may also differ from nonrespondents based not on will (interest and motivation) but based on ability. Thus, if some respondents have abilities that allow them to plan and organize their (work) life better than others, it might be easier for them to find the time to respond to a survey.

Naturally, these insights from the literature on nonresponse are important for a general perspective. But concerning public management research, they also shed light on an often neglected question: How representative are respondents in leadership surveys in relation to the general population of managers? We will now discuss how nonresponse bias may develop in leadership studies.

LEADERSHIP AND NONRESPONSE BIAS

Leadership in public organizations has become a widely studied topic in the public management literature over the past decades (Vogel and Masal 2015; Van Wart 2013). Although leadership has traditionally been regarded as less important in public organizations (Rainey 2014), more recent studies have linked leadership with several desirable outcomes such as public service motivation (Park and Rainey 2008; Wright, Moynihan, and Pandey 2012), performance information use (Kroll and Vogel 2014), work quality (Oberfield 2012), and ultimately performance (Bellé 2014; Jacobsen and Andersen 2015; Jacobsen et al. 2021). Furthermore, the interest in leadership as a research topic seems to rise (Knies, Jacobsen, and Tummers 2016).

Leadership studies also face several challenges regarding data collection. To begin with, surveying subordinates about their supervisor's leadership behavior requires a mechanism to match follower responses to the responses of their supervisors (Vogel 2018). Additionally, subordinates might have different interest in and motivation to participate in such surveys, raising the issue of nonresponse bias. Furthermore, the same nonresponse issues arise when asking leaders to self-assess their own leadership behavior. Such self-assessments also tend to be flawed because people have difficulties assessing their own behavior (Vogel and Kroll 2019; Jacobsen and Andersen 2015; Fleenor et al. 2010). Hence, leadership researchers avoid self-assessments. However, they often must rely on other information provided by leaders, such as supervisor-assessed performance (e.g., Potipiroon and Faerman 2016; Wright, Hassan, and Christensen 2017) or self-assessments of leaders' attitudes, values, or beliefs (e.g., Jensen, Andersen, and Jacobsen

2019). Although such measures are very important and increasingly used by scholars, they also depend on leaders' willingness to allocate time to replying to a questionnaire, raising the issue of nonresponse bias.

Therefore, nonresponse bias is a concern in leadership studies regardless of whether it stems from the response patterns of subordinates or leaders. In this article, we focus on potential nonresponse bias in surveys of leaders. We do so because of the importance leaders' self-reports, as noted earlier, and because a number of prior leadership studies offer the opportunity to compare information on responding leaders with those of nonresponding leaders. Specifically, studies that include surveys of both leaders and subordinates allow leaders' self-reports to be compared with the reporting of leadership behavior by subordinates. This gives us the rare opportunity to learn something about the differences between responding and nonresponding leaders in surveys.

The literature on leadership in public organizations already covers a wide understanding of leadership (Van Wart 2013), but two broadly studied leadership strategies are transformational and transactional leadership (Knies, Jacobsen, and Tummers 2016). Transformational and transactional leadership are both described as goal-oriented strategies, but they differ in terms of underlying logic and approaches (Bass 1985; Jensen et al. 2019). Thus, transformational leaders seek to develop, share and sustain a vision that encourages employees to transcend their self-interest and achieve organizational goals. In contrast, transactional leaders use contingent rewards and

sanctions to direct employee self-interest toward achieving organizational goals (Jacobsen and Andersen 2015, 832).

Recent developments in public sector leadership research have led to the study of additional leadership approaches. One approach adopted from private sector research is servant leadership. Servant leaders enable their followers by putting their followers' needs at the center of their work and seeing themselves as a servant of their followers (van Dierendonck 2011). They “strive selflessly to assist others before themselves and encourage their followers to do the same” (Schwarz et al. 2016, 1025). Another leadership approach adapted from private sector research is authentic leadership. The approach emphasizes that a leader should be “true to him/herself” and “confident, hopeful, optimistic, resilient, moral/ethical, future-oriented, and gives priority to developing associates to be leaders” (Luthans and Avolio 2003, 243). The approach is commonly conceptualized as a four-dimensional construct, with the dimensions of self-awareness, relational transparency, balanced processing, and internalized moral perspective (Neider and Schriesheim 2011).

In contrast to leadership approaches adopted from private sector research, Tummers and Knies (2016) developed the public leadership construct as a specific leadership approach for the public sector. It focuses on the aspects of leadership that are specifically public (Vogel and Masal 2015) and specifies four distinct roles of public leaders: accountability leadership, rule-following leadership, political loyalty leadership, and network governance leadership. Tummers and Knies (2016) show that these roles are

positively correlated with leadership effectiveness as well as followers' organizational commitment, work engagement, organizational citizenship behavior, and job satisfaction.

We expect that the measurement of all these leadership approaches may be susceptible to nonresponse bias. In particular, leaders with the will and ability to exert leadership are also likely to be associated with a higher propensity to respond to surveys about leadership and organizational affairs. Active leaders have in common that they care for their organizations, which might also include the willingness to participate in surveys that are also intended to learn about and improve their organizations. Therefore, we hypothesize that responding and nonresponding leaders will be evaluated differently by their followers as follows:

H1: Employees of responding leaders will report higher values of leadership behavior than employees of nonresponding leaders.

RESEARCH DESIGN AND DATA

This article aims at a broad assessment of potential differences between responding and nonresponding leaders in the public sector. Accordingly, we identified published research that uses matched data of leaders and their followers (Vogel 2018) and asked the authors to share their data. In addition, we published a call for datasets in three public administration mailing lists (International Public Management Network, Public and Nonprofit Division of the Academy of Management, Experimental and Behavioral Public Administration). In this way, we were able to obtain six datasets suitable for our

analysis and, in the process, demonstrate the potential of collaboration among public administration researchers and the opportunities that come with open science practices (Perry 2017; Nosek et al. 2015). We thank *Stephan Dorsman, Artur Reuber, Bram Steijn, Joris Van der Voet, and Rick Vogel*, for sharing their data.

The datasets used in the analyses are summarized in table 1. They include responses from 16,531 followers and 1,164 leaders, originate from three different countries, and provide data on five different leadership approaches. Leader response rates varied from 42.2 % to 97.2 %.

These data offer the opportunity to compare followers' assessments of leaders who did respond in the survey with the assessments of leaders who did not respond. In this way, we have the rare opportunity to compare the leadership approaches of responding and nonresponding leaders and assess whether they differ significantly. Any observed difference in the assessment of responding and nonresponding leaders provides an indicator of nonresponse bias.

Table 1: Analyzed datasets

Surveyed organizations	Country	Followers (response rate)	Leaders (response rate)	Leadership approaches	Reference
127 high schools	Denmark	4,255 teachers (32.6 %)	99 principals (67.3 %)	Transformational, Transactional (rewarding, sanctioning)	Jacobsen and Andersen 2015
2 state agencies, 1 county	Germany	471 public servants (34.5 %)	64 street-level managers (57.1 %)	Transformational	Vogel and Kroll 2019
Primary and secondary schools, daycare centers, tax offices	Denmark	6,365 employees (45.6 %)	375 managers (97.2 %)	Transformational, Transactional (nonpecuniary rewards, pecuniary rewards, sanctions)	Nielsen, Boye, and Holten 2019
146 local welfare teams	Netherlands	1,358 welfare professionals (53.2 %)	146 team leaders (92.4 %)	Transformational	Van der Voet and Steijn 2020
7 city district offices	Germany	2,274 public servants (30.4 %)	368 managers (42.2 %)	Public leadership, Authentic leadership	Vogel, Reuber, and Vogel 2020
Employee Insurance Agency	Netherlands	1,808 public servants (31.0 %)	112 managers (46.7 %)	Servant leadership, public leadership (rule-following leadership)	Dorsman 2017

Note: Samples sizes and response rates are based on the observations that could be used for the analysis.

One challenge in using this kind of strategy to assess nonresponse bias in surveys of leaders is that, of course, there is also nonresponse in the matched survey of followers. Table 1 shows that the degree of nonresponse by followers in our data sets varies between 50.9 % and 69.6 %. For four datasets, we were able to directly compare the rate of follower nonresponse for both responding and nonresponding leaders. Table 2

summarizes the results and shows that the difference is quite small for three datasets but more substantial for the remaining one.

Table 2: Followers' response rates separated by the response status of the supervisor

Dataset	Responding supervisor	Non-responding supervisor
Danish high schools	43.8%	41.4%
Dutch unemployment agency	34.4%	35.6%
Danish schools, daycare centers, and tax offices	45.7%	42.3%
German city district offices	34.2%	27.8%

One might argue that follower nonresponse can lead to nonresponse bias in the subsequent analyses of leader nonresponse. However, this would only be the case if nonresponding followers of responding leaders differ from those of nonresponding leaders. To test if this is the case, we compared the two groups with regard to age, gender, and (where available) education of the followers. Of the 13 tested differences on these variables, only two yielded significant results. For the German state agencies and the Dutch unemployment agency, followers differed significantly in their age. Although such demographics provide only a limited view on potential differences between responding and nonresponding followers, they are the best gauge we have at hand to probe this issue. Still, these results are encouraging in that they do not indicate the likelihood of much nonresponse bias in the surveys of followers. In addition, it is reasonable to assume that the factors and motivations driving nonresponse among followers may well be different from (and thus somewhat independent of) the factors and motivations driving leader nonresponse. If so, any nonresponse bias in the survey of followers would not likely

influence the assessment of nonresponse bias in the survey of leaders. However, such confounding remains a possibility that we will consider and return to in our interpretation of results.

MEASUREMENT

The datasets reflect the diverse perspectives that can be taken to study leadership, its antecedents, and its effects and therefore includes various leadership approaches like transformational, transactional, public, servant, and authentic leadership. It also reflects the different measures used to assess the same leadership concept. Appendix A lists all operationalizations used in each dataset in detail. Transformational leadership, for example, is measured by items developed by Podsakoff, MacKenzie, and Bommer (1996), the Federal Employee Viewpoint Survey team (Trottier, Van Wart, and Wang 2008), Wright, Moynihan, and Pandey (2012), and Jensen et al. (2019). This might be seen as a limitation of our analyses, but we would argue that this mix of measures in realistically reflects the variety of different approaches to leadership in the public management literature.

ANALYTICAL PROCEDURE

Given that the data are clustered into organizations or teams, we use hierarchical linear modeling instead of ordinary least squares regression or t-tests to assess whether followers perceive responding managers to exert more leadership. Ignoring the clustered

structure of the data would result in estimations overweighting bigger organizations or teams because every response is treated independently. Therefore, we use a random intercept model (Garson 2013) with the followers' assessment of the respective leadership behavior as the dependent variables and a dummy indicating if the leader responded as the only independent variable. If the dummy's coefficient is significantly different from zero, we can infer that followers of a responding manager report different leadership behavior than followers of a nonresponding manager. In order to make the results comparable between the different datasets, we standardized the dependent variable (i.e., leadership behavior).

In this context, it is important to note that an insignificant finding does not indicate that there is no difference between responding and nonresponding managers, i.e., the data does not necessarily support the null hypothesis (Lakens et al. 2020). Rather, null hypothesis significance testing (NHST) with p values indicates how likely such an effect (or a more extreme one) is if the null hypothesis is true (Wasserstein and Lazar 2016). Thus, an insignificant p value does not mean that the null hypothesis of no difference is true. It is possible that the true difference is zero, of course, but it could also be the case that the study does not have enough power to detect the true effect.

Therefore, it is necessary to use alternative statistical procedures to test if the null hypothesis is supported by the data, i.e., to differentiate between data that support the null hypothesis and data that neither support the null hypothesis nor the alternative hypothesis of a non-zero effect (Harms and Lakens 2018). To do so, we use two approaches that are based on Bayesian statistics (see Kruschke and Liddell 2018a and Gill and Witko 2013

for a primer on Bayesian statistics): Bayes factors (Wagenmakers et al. 2010) and the so-called region of practical equivalence (ROPE) (Kruschke and Liddell 2018b).

Bayesian statistics are based on three elements: A model of the probability of data before the data is collected (the so-called prior), a likelihood function to calculate a posterior distribution based on the data, and the posterior after the data is collected. The posterior is “our belief about different parameter values [...] after having seen the data” (Harms and Lakens 2018, 386). The posterior, therefore, expresses how likely certain values are, given the data. In our case, the parameter of interest is the difference in reported leadership behavior between responding and nonresponding leaders. The prior describes how likely we believe different values of this parameter to be before we conducted our analyses. We used a normal distribution with a mean of zero and a standard deviation of one—a so-called generic weakly informative prior (Stan Development Team 2019)—to describe this prior belief. The posterior describes our beliefs of the probability of certain values after we updated the prior with our data.

Using the prior and posterior distribution of our parameter of interest, we can calculate how much more or less likely a null effect is than a non-zero effect when we consider the collected data. This ratio is called the Bayes factor (BF). A Bayes factor can be expressed in favor of the null effect (BF_{01}) or a non-zero effect (BF_{10}). A BF_{01} of 5.0 means that a null effect is five times more likely given the data we collected than a non-zero effect. This way, the Bayes factor expresses how our beliefs changed after we analyzed the data and allows us to judge whether the data allow for a conclusion of the

absence of a difference between responding and nonresponding managers (Harms and Lakens 2018).

An alternative approach is used for the region of practical equivalence (ROPE). For the ROPE, the distribution of parameter values in the posterior distribution is described by a Highest Density Interval (HDI). A 90 % HDI specifies an interval that contains 90 % of the values of a distribution. The values within the 90 % HDI are the 90 % most credible values of the distribution. The ROPE procedure compares this HDI with a pre-specified region of practical equivalence, which is an interval researchers perceive to be so small that it is essentially equivalent to zero. In our case, we decided that differences between responding and nonresponding managers that are smaller than 0.2 standard deviations of the leadership behavior of interest are too small to take into consideration. This decision is based on common perceptions that standardized mean differences (i.e., Cohen's d) are small if they are 0.2 or more (Cohen 1988). Comparing the HDI with the ROPE, we can report what proportion of credible values lies inside the region of practical equivalence. If all credible values fall inside the ROPE, the parameter of interest is practically not distinguishable from zero.

The following analyses are conducted using R version 4.0.2 (R Core Team 2020) with the *nlme* package (Jose Pinheiro et al. 2018) for estimating the multilevel regression models, the *brms* package (Bürkner 2017) for estimating the Bayesian models, the *bayestestR* package (Makowski, Ben-Shachar, and Lüdtke 2019) for conducting the ROPE procedure, and the *metaphor* package (Viechtbauer 2010) for estimating an internal meta-analysis.

Table 3: Descriptive statistics

Variable	n	Mean	St. Dev.	Min	Max
<i>Danish high schools</i>					
Transformational Leadership	4,255	2.999	1.002	1.000	5.000
Rewarding Leadership	4,255	2.479	1.028	1.000	5.000
Sanctioning Leadership	4,255	2.349	0.937	1.000	5.000
Principal response (1 = yes)	4,255	0.646	0.480	0	1
<i>German state agencies</i>					
Transformational Leadership	471	2.802	1.038	1.000	5.000
Manager response (1 = yes)	471	0.466	0.501	0	1
<i>Danish schools, daycare centers, and tax offices</i>					
Transformational Leadership	6,365	3.805	0.869	1.000	5.000
Nonpecuniary rewards	6,365	3.502	1.068	1.000	5.000
Pecuniary rewards	6,365	2.523	0.942	1.000	5.000
Sanctioning Leadership	6,365	2.897	0.834	1.000	5.000
Manager response (1 = yes)	6,365	0.971	0.168	0	1
<i>Dutch local welfare teams</i>					
Transformational Leadership	1,358	3.672	0.845	1.000	5.000
Manager response (1 = yes)	1,358	0.939	0.239	0	1
<i>German city district offices</i>					
Public Leadership	2,274	3.355	0.742	1.000	5.000
Authentic Leadership	2,274	3.384	0.855	1.000	5.000
Manager response (1 = yes)	2,274	0.458	0.499	0	1
<i>Dutch unemployment insurance agency</i>					
Servant leadership	1,602	5.073	0.870	1.821	7.000
Rule-following leadership	1,602	4.925	1.401	1.000	7.000
Manager response (1 = yes)	1,602	0.404	0.492	0	1

RESULTS

Table 3 gives an overview of the descriptive statistics of all variables in the datasets.

In the next step, we look at the differences in leadership behavior between responding and nonresponding managers. Table 4 summarizes the results of 14 hierarchical linear models testing group differences. Appendix B gives detailed results of all estimated models.

Overall, the random-intercept multilevel models revealed only one statistically significant difference between responding and nonresponding managers. In the dataset collected from employees of six German city district offices, subordinates of responding managers report 0.127 standard deviations higher values of public leadership than their colleagues working for a nonresponding manager. This corresponds to a difference of 0.094 points on a five-point Likert scale ($p = .020$). The other effects range from -0.278 (transformational leadership in the Danish schools, daycare centers, and tax offices dataset) to 0.229 (transformational leadership in the German state agencies dataset) standard deviations, although both of them are not statistically significant.

Because of the insignificant results, we need to clarify if the data supports the notion that responding and nonresponding public managers' leadership behavior is assessed equally by their followers or if additional data is necessary. To answer this question, we first look at the Bayes factors. All of them are greater than one, indicating that the data is more in favor of a null effect than a non-zero effect. However, the Bayes factors differ in how strongly they favor the null effect. The values range from about 1.4

(public leadership in the German city district offices dataset) to 15.8 times more likely (rewarding leadership in the Danish high schools dataset). Conventionally, Bayes factors between one and three are considered anecdotal evidence, between three and ten as moderate evidence, and above ten as strong evidence (Wagenmakers et al. 2018). In Table 4, there are four Bayes factors between 1 and 3 (anecdotal evidence), five between 3 and 10 (moderate evidence), and four above 10 (strong evidence). Overall, the evidence for the absence of differences in leadership assessment between responding and nonresponding managers seems moderate.

As an additional approach to assess whether the data allow us to conclude that there is no difference between responding and nonresponding managers, we look at the region of practical equivalence. The values in the last column of Table 4 indicate what proportion of the 90 % highest density interval (HDI) lays within a range of -0.2 to 0.2 (i.e., the region of practical equivalence; ROPE). We can see that these values vary between 32% and 100%. For five effects, the full HDI falls within the ROPE. Overall, the ROPE approach also supports the notion that responding and nonresponding managers are, in practical terms, not evaluated differently.

In order to summarize our results, we conducted an internal meta-analysis (Maner 2014) to estimate an overall effect size of the difference between responding and nonresponding managers. The meta-analysis uses a mixed-effects model that takes into account that the effects are clustered in datasets. The results are displayed in Figure 1. The model estimates an overall effect size of 0.054 standard deviations which is not significantly different from zero ($p = .057$, 95 % $CI = -0.002, 0.109$). Using the weights

determined by the meta-analytical model, we can calculate an overall Bayes factor of 7.422 in favor of a null effect. As noted earlier, this indicates moderate evidence of no difference.

Table 4: Standardized differences between responding and nonresponding leaders

Study	Leadership approach	Coefficient (SE)	BF₀₁	ROPE
Danish high schools	Transformational Leadership	0.142 (0.079)	2.562	80.14 %
	Rewarding Leadership	0.005 (0.063)	15.772	100.00 %
	Sanctioning Leadership	-0.044 (0.074)	11.312	100.00 %
German state agencies	Transformational Leadership	0.229 (0.130)	1.689	41.13 %
Danish schools, daycare centers, and tax offices	Transformational Leadership	-0.278 (0.176)	1.786	32.28 %
	Nonpecuniary rewards	-0.143 (0.170)	4.266	66.10 %
	Pecuniary rewards	0.133 (0.163)	4.509	69.19 %
	Sanctioning Leadership	0.155 (0.137)	3.972	64.99 %
Dutch local welfare teams	Transformational Leadership	-0.056 (0.177)	5.567	81.62 %
German city district offices	Public Leadership	0.127 (0.054)*	1.365	96.51 %
	Authentic Leadership	0.099 (0.056)	3.942	100.00 %
Dutch unemployment insurance agency	Servant leadership	-0.016 (0.072)	13.331	100.00 %
	Rule-following leadership	0.037 (0.072)	12.276	100.00 %

Notes: Coefficient = Coefficients of hierarchical linear models with restricted maximum likelihood estimator; SE = Standard errors; BF₀₁ = Bayes Factor of a point-null hypothesis against the estimated model using the Savage-Dickey density ratio method; ROPE = Region of practical equivalence (proportion of values within the 90 % highest density interval of the posterior distribution within the region of practical equivalence (-0.2 to 0.2)); * $p < 0.05$, ** $p < 0.01$

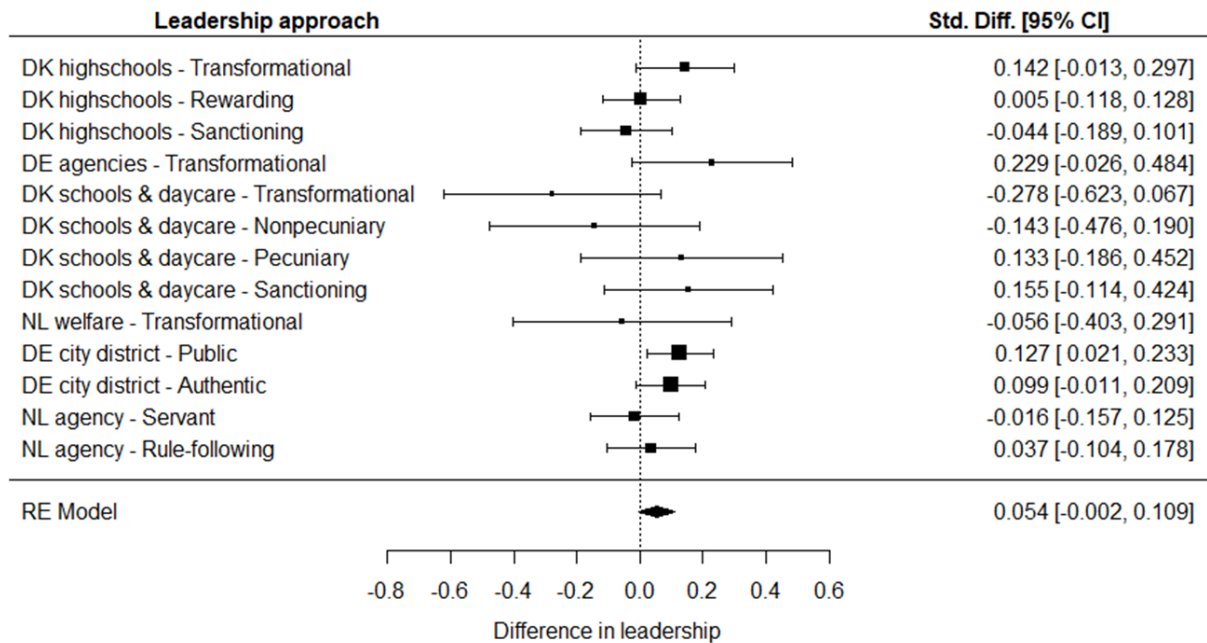


Figure 1: Results of mixed-effects meta-analysis using restricted maximum likelihood estimation and taking clustering in datasets into account. Each row represents the estimation of the standardized difference between responding and nonresponding managers. Squares represent the estimated standardized difference; bars represent 95 % confidence intervals of the estimated difference. Last row reports the estimation of the meta-analytical effect sizes

DISCUSSION

Nonresponse is rarely investigated across research areas because of the difficulty of obtaining data on nonrespondents. For public leadership and public management research, moreover, it should be a concern to assess whether samples are representative of the larger population of public managers they are drawn from. Indeed, many studies of leadership in public management depend on information gathered from surveys of

leaders. But if respondents differ from nonrespondents in their leadership behavior as well as the outcomes under investigation, the results of such studies are likely to be biased. Our investigation has shown that nonresponse bias can be identified in at least one of six datasets, which allowed the comparison of surveys of both leaders and followers. However, additional analyses indicate that, overall, the extent of nonresponse bias in the assessed datasets is indistinguishable from a null effect. Still, the degree of certainty with which we can reject a nonresponse bias differs between datasets and is overall only moderate. In addition, nonresponse in the survey of followers may confound the analysis of nonresponse bias in the survey of leaders. Hence, although our results suggest only limited nonresponse bias in leadership surveys in public management research, they do not imply that researchers should be unconcerned about the problem.

Importantly, our study represents a broad sample of leadership constructs, countries, and organizations studied in public management research. The datasets also vary in leaders' response rate. Although we deliberately sought such a broad perspective, this does make it somewhat difficult to assess what specifically might drive different degrees of potential nonresponse bias. However, we do not see a clear pattern of how leaders' response rate influence measures of leadership behavior. For example, based on our results, it cannot be stated that higher response rates lead to lower differences between respondents' and nonrespondents' answers. Additional research is needed to test if certain cultures, leadership approaches, or public service settings are more prone to nonresponse bias. The results of our study call for more attention to the causal inferences drawn from studies that involve some sort of measures obtained from leaders (e.g., self-

assessments of leadership, attitudes, or supervisor-provided follower information such as performance).

What can researchers and practitioners learn from these results? First, this study underlines that researchers should be careful when using measures that are obtained from leaders. Although nonresponse bias seems not to be a widespread issue, it cannot be entirely dismissed. Second, researchers should assess potential nonresponse bias in their own data. Hence, because responding and nonresponding managers might differ in some aspects of leadership, we would caution against organizing data collection in a two-step procedure where follower data is only collected for those organizations or units where the manager already participated. This procedure leaves researchers with potentially flawed data, importantly, and no opportunity to check for nonresponse bias.

In addition, our study clearly confirms methodological knowledge that is well known to all researchers and practitioners but often ignored: nonresponse matters, and we should put in as much effort as possible to avoid it (Bethlehem, Cobben, and Schouten 2011). What can researchers do to avoid nonresponse bias? Based on our results, we can provide a few answers. Besides the obvious measures like reminding nonresponding participants of the study, ensure anonymity, and gain top management support, survey construction is a point to consider. On the one hand, many leaders are interested in leadership, and response rates may be increased through clear communication of the content to potential respondents. Despite our finding of limited bias, this is helpful since nonresponse bias is the difference between responding and nonresponding participants multiplied by the nonresponse rate. Hence, it is still true that the impact of nonresponse

bias is *ceteris paribus* the smaller, the higher the response rate is. However, everything may not be equal because the same interest mechanism can appeal to some leaders more than others. If interested responders also exert more leadership, communicating too much about content may increase nonresponse bias. Studies have also attempted to lure respondents into providing answers through incentive schemes (e.g., Pedersen and Nielsen 2016), but if this appeals to more extrinsically motivated leaders, the increased response rate may incur the cost of nonresponse bias. We still have very little knowledge about these effects, so future studies should look into these important questions.

We must also note some limitations of our study. First, we have based our research on only six datasets containing data on only some aspects of leadership. Although this brings the knowledge on nonresponse bias forward in the field of public management, it represents just the beginning of an important research agenda on nonresponse bias that should be extended to other leadership strategies, types of respondents, public service areas, and countries. We presume other researchers might have data comparable to the ones we used here, with surveys of both leaders and followers. We encourage such researchers to report results on differences between responding and nonresponding leaders, as well as the effects of nonresponse on important outcomes. Furthermore, it would be extremely useful if future studies move beyond just reporting response rates to investigating methods to limit nonresponse bias. Second, our leadership measures are based on employee-rated leadership and, just like opening one Russian Matryoshka doll leads to the next doll, we also need to consider the potential nonresponse bias in employee surveys. We tried to rule out the possibility that follower

nonresponse bias was a serious issue in our data, but the tools available to us – although providing encouraging results – offer only limited evidence for the absence of follower nonresponse bias. Third, one of the analyzed datasets – the Danish schools, daycare centers, and tax offices data – includes only those leaders who initially signed up to participate in a larger research project. This might affect survey results about their behavior as well as the behavior of their followers. We tried to limit this impact by using the second wave of the survey, but there might still be an effect. However, we do not see stark contrasts to the other datasets analyzed.

This article provides a unique view into potential nonresponse bias in leadership research. It finds only limited evidence for nonresponse bias in leadership research, at least with regard to the self-reported behavior of leaders. Although this is an encouraging finding for leadership researchers, we urge researchers to not ignore (potential) nonresponse bias and realize that multisource data also provides multiple sources for nonresponse bias.

REFERENCES

- Adigüzel, Feray, and Michel Wedel. 2008. Split Questionnaire Design for Massive Surveys. *Journal of Marketing Research* 45 (5): 608–17. doi: 10.1509/jmkr.45.5.608.
- Angrist, Joshua D., and Jörn-Steffen Pischke. 2010. The Credibility Revolution in Empirical Economics: How Better Research Design is Taking the Con out of Econometrics. *Journal of Economic Perspectives* 24 (2): 3–30. doi: 10.1257/jep.24.2.3.
- . 2014. *Mastering Metrics: The Path from Cause to Effect*. Princeton: Princeton University Press.
- Armstrong, J. S., and Terry S. Overton. 1977. Estimating Nonresponse Bias in Mail Surveys. *Journal of Marketing Research* 14 (3): 396. doi: 10.2307/3150783.
- Bass, B. M. 1985. *Leadership and Performance Beyond Expectations*. New York: Free Press.
- Bellé, Nicola. 2014. Leading to Make a Difference: A Field Experiment on the Performance Effects of Transformational Leadership, Perceived Social Impact, and Public Service Motivation. *Journal of Public Administration Research and Theory* 24 (1): 109–36. doi: 10.1093/jopart/mut033.
- Bethlehem, J., F. Cobben, and B. Schouten. 2011. *Handbook of Nonresponse in Household Surveys*. Hoboken, NJ: Wiley.
- Bickart, Barbara, and David Schmittlein. 1999. The Distribution of Survey Contact and Participation in the United States: Constructing a Survey-Based Estimate. *Journal of Marketing Research* 36 (2): 286–294. doi: 10.2307/3152100.
- Bürkner, Paul-Christian. 2017. brms: An R Package for Bayesian Multilevel Models Using Stan. *Journal of Statistical Software* 80 (1). doi: 10.18637/jss.v080.i01.
- Cohen, Jacob. 1988. *Statistical power analysis for the behavioral sciences*. 2nd ed. Hillsdale, NJ: Erlbaum.
- de Leeuw, Edith, and Wim de Heer. 2002. Trends in household survey nonresponse: A longitudinal and international comparison. In *Survey Nonresponse*, ed. Robert M. Groves, D. A. Dillman, J. L. Eltinge, and R. J. A. Little, 41–54. New York: Wiley.
- Dorsman, Stephan J. 2017. Leadership and innovation in the public sector: Studying the organizational, team and individual level. PhD thesis, Erasmus University Rotterdam. <http://hdl.handle.net/1765/100429>.
- Edwards, Philip J., Ian Roberts, Mike J. Clarke, Carolyn Diguiseppi, Reinhard Wentz, Irene Kwan, Rachel Cooper, Lambert M. Felix, and Sarah Pratap. 2009. Methods to increase response to postal and electronic questionnaires. *The Cochrane Database of Systematic Reviews* (3). doi: 10.1002/14651858.MR000008.pub4.

- Fleenor, John W., James W. Smither, Leanne E. Atwater, Phillip W. Braddy, and Rachel E. Sturm. 2010. Self–other rating agreement in leadership: A review. *The Leadership Quarterly* 21 (6): 1005–34. doi: 10.1016/j.leaqua.2010.10.006.
- Garson, G. D. 2013. Fundamentals of Hierarchical Linear and Multilevel Modeling. In *Hierarchical linear modeling: Guide and applications*, ed. G. D. Garson, 3–25. Thousand Oaks, Calif. Sage.
- Gill, Jeff, and Christopher Witko. 2013. Bayesian Analytical Methods: A Methodological Prescription for Public Administration. *Journal of Public Administration Research and Theory* 23 (2): 457–94. doi: 10.1093/jopart/mus091.
- Groves, Robert M. 2006. Nonresponse Rates and Nonresponse Bias in Household Surveys. *Public Opinion Quarterly* 70 (5): 646–75. doi: 10.1093/poq/nfl033.
- Groves, Robert M., D. A. Dillman, J. L. Eltinge, and R. J. A. Little, eds. 2002. *Survey Nonresponse*. New York: Wiley.
- Groves, Robert M., and E. Peytcheva. 2008. The Impact of Nonresponse Rates on Nonresponse Bias: A Meta-Analysis. *Public Opinion Quarterly* 72 (2): 167–89. doi: 10.1093/poq/nfn011.
- Groves, Robert M., Eleanor Singer, and Amy Corning. 2000. Leverage-Saliency Theory of Survey Participation: Description and an Illustration. *The Public Opinion Quarterly* 64 (3): 299–308.
- Harms, Christopher, and Daniël Lakens. 2018. Making 'null effects' informative: statistical techniques and inferential frameworks. *Journal of Clinical and Translational Research* 3 (S2): 382–93. doi: 10.18053/jctres.03.2017S2.007.
- Hartog, Deanne N. D., Jaap J. van Muijen, and Paul L. Koopman. 1997. Transactional versus transformational leadership: An analysis of the MLQ. *Journal of Occupational & Organizational Psychology* 70 (1): 19–34.
- Jacobsen, Christian B., and Lotte B. Andersen. 2015. Is Leadership in the Eye of the Beholder? A Study of Intended and Perceived Leadership Practices and Organizational Performance. *Public Administration Review* 75 (6): 829–41. doi: 10.1111/puar.12380.
- Jacobsen, Christian B., Lotte B. Andersen, Anne Bøllingtoft, and Tine L. M. Eriksen. 2021. Can Leadership Training Improve Organizational Effectiveness? Evidence from a Randomized Field Experiment on Transformational and Transactional Leadership. *Public Administration Review*. doi: 10.1111/puar.13356.
- Jensen, U. T., Lotte B. Andersen, L. L. Bro, A. Bollingtoft, T. L. M. Eriksen, A.-L. Holten, Christian B. Jacobsen et al. 2019. Conceptualizing and Measuring Transformational and Transactional Leadership. *Administration & Society* 51 (1): 3–33. doi: 10.1177/0095399716667157.

- Jensen, Ulrich T., Lotte B. Andersen, and Christian B. Jacobsen. 2019. Only When We Agree! How Value Congruence Moderates the Impact of Goal-Oriented Leadership on Public Service Motivation. *Public Administration Review* 79 (1): 12–24. doi: 10.1111/puar.13008.
- Jose Pinheiro, Douglas Bates, Saikat DebRoy, Deepayan Sarkar, and R Core Team. 2018. nlme: Linear and Nonlinear Mixed Effects Models. <https://cran.r-project.org/package=nlme>.
- Knies, Eva, Christian B. Jacobsen, and Lars G. Tummars. 2016. Leadership and organizational performance: State of the art and research agenda. In *The Routledge companion to leadership*, ed. John Storey, Jean Hartley, Jean-Louis Denis, Paul 't Hart, and David Ulrich, 404–18. New York: Routledge.
- Kreuter, F., G. Muller, and M. Trappmann. 2010. Nonresponse and Measurement Error in Employment Research: Making Use of Administrative Data. *Public Opinion Quarterly* 74 (5): 880–906. doi: 10.1093/poq/nfq060.
- Kroll, Alexander, and Dominik Vogel. 2014. The PSM-leadership fit: A model of performance information use. *Public Administration* 92 (4): 974–91. doi: 10.1111/padm.12014.
- Kruschke, John K., and Torrin M. Liddell. 2018a. Bayesian data analysis for newcomers. *Psychonomic Bulletin & Review* 25 (1): 155–77. doi: 10.3758/s13423-017-1272-1.
- . 2018b. The Bayesian New Statistics: Hypothesis testing, estimation, meta-analysis, and power analysis from a Bayesian perspective. *Psychonomic Bulletin & Review* 25 (1): 178–206. doi: 10.3758/s13423-016-1221-4.
- Lahaut, Viviënne M. H. C. J., Harrie A. M. Jansen, Dike van de Mheen, and Henk F. L. Garretsen. 2002. Non-Response Bias in a Sample Survey on Alcohol Consumption. *Alcohol and Alcoholism* 37 (3): 256–60. doi: 10.1093/alcalc/37.3.256.
- Lakens, Daniël, Neil McLatchie, Peder M. Isager, Anne M. Scheel, and Zoltan Dienes. 2020. Improving Inferences about Null Effects with Bayes Factors and Equivalence Tests. *Journals of Gerontology: Psychological Sciences* 75 (1): 45–57. doi: 10.1093/geronb/gby065.
- Lambert, Douglas M., and Thomas C. Harrington. 1990. Measuring nonresponse bias in customer service mail surveys. *Journal of Business Logistics* 11 (2): 5–25.
- Luthans, Fred, and Bruce J. Avolio. 2003. Authentic leadership: A positive development approach. In *Positive Organizational Scholarship: Foundations of a New Discipline*, ed. K. Cameron and Jane E. Dutton, 241–61. San Francisco, CA: Berrett-Koehler.
- Makowski, Dominique, Mattan S. Ben-Shachar, and Daniel Lüdtke. 2019. Understand and Describe Bayesian Models and Posterior Distributions using bayestestR. *CRAN*. doi: 10.5281/zenodo.2556486.

- Maner, Jon K. 2014. Let's Put Our Money Where Our Mouth Is: If Authors Are to Change Their Ways, Reviewers (and Editors) Must Change With Them. *Perspectives on Psychological Science* 9 (3): 343–51. doi: 10.1177/1745691614528215.
- Manfreda, Katja L., Michael Bosnjak, Jernej Berzelak, Iris Haas, and Vasja Vehovar. 2008. Web surveys versus other survey modes: a meta-analysis comparing response rates. *International Journal of Market Research* 50 (1): 79–104.
- Meier, K. J., and Laurence J. O'Toole. 2011. Comparing Public and Private Management: Theoretical Expectations. *Journal of Public Administration Research and Theory* 21 (Supplement 3): i283–i299. doi: 10.1093/jopart/mur027.
- Neider, Linda L., and Chester A. Schriesheim. 2011. The Authentic Leadership Inventory (ALI): Development and empirical tests. *The Leadership Quarterly* 22 (6): 1146–64. doi: 10.1016/j.leaqua.2011.09.008.
- Nielsen, Poul A., Stefan Boye, and Ann-Louise Holten. 2019. Are Transformational and Transactional Types of Leadership Compatible? A Two-Wave Study of Employee Motivation. *Public Administration* 97 (2): 413–28. doi: 10.1111/padm.12574.
- Nosek, B. A., G. Alter, G. C. Banks, D. Borsboom, S. D. Bowman, S. J. Breckler, S. Buck et al. 2015. Promoting an open research culture. *Science* 348 (6242): 1422–25. doi: 10.1126/science.aab2374.
- Oberfield, Zachary W. 2012. Motivation, Change, and Stability: Findings From an Urban Police Department. *American Review of Public Administration* 44 (2): 210–32. doi: 10.1177/0275074012461297.
- . 2014. Public Management in Time: A Longitudinal Examination of the Full Range of Leadership Theory. *Journal of Public Administration Research and Theory* 24 (2): 407–29. doi: 10.1093/jopart/mus060.
- Park, Sung M., and Hal G. Rainey. 2008. Leadership and Public Service Motivation in U.S. Federal Agencies. *International Public Management Journal* 11 (1): 109–42. doi: 10.1080/10967490801887954.
- Pedersen, M. J., and C. V. Nielsen. 2016. Improving Survey Response Rates in Online Panels: Effects of Low-Cost Incentives and Cost-Free Text Appeal Interventions. *Social Science Computer Review* 34 (2): 229–43. doi: 10.1177/0894439314563916.
- Perry, James L. 2017. Practicing What We Preach! Public Administration Review Promotes Transparency and Openness. *Public Administration Review* 77 (1): 5–6. doi: 10.1111/puar.12705.
- Podsakoff, Philip M., Scott B. MacKenzie, and William H. Bommer. 1996. Transformational Leader Behaviors and Substitutes for Leadership as Determinants of Employee Satisfaction, Commitment, Trust, and Organizational Citizenship Behaviors. *Journal of Management* 22 (2): 259–98. doi: 10.1177/014920639602200204.

- Potipiroon, Wisanupong, and Sue Faerman. 2016. What Difference Do Ethical Leaders Make? Exploring the Mediating Role of Interpersonal Justice and the Moderating Role of Public Service Motivation. *International Public Management Journal* 19 (2): 171–207. doi: 10.1080/10967494.2016.1141813.
- R Core Team. 2020. *R: A Language and Environment for Statistical Computing*. Vienna, Austria. <https://www.r-project.org/>.
- Rainey, Hal G. 2014. *Understanding and Managing Public Organizations*. 5th ed. New York, NY: Jossey-Bass.
- Schwarz, Gary, Alexander Newman, Brian Cooper, and Nathan Eva. 2016. Servant leadership and follower job performance: The mediating effect of public service motivation. *Public Administration* 94 (4): 1025–41. doi: 10.1111/padm.12266.
- Stan Development Team. 2019. Prior Choice Recommendations. Accessed August 15, 2019. <https://github.com/stan-dev/stan/wiki/Prior-Choice-Recommendations>.
- Trottier, Tracey, Montgomery Van Wart, and XiaoHu Wang. 2008. Examining the Nature and Significance of Leadership in Government Organizations. *Public Administration Review* 68 (2): 319–33. doi: 10.1111/j.1540-6210.2007.00865.x.
- Tummers, Lars G., and Eva Knies. 2016. Measuring Public Leadership: Developing Scales for Four Key Public Leadership Roles. *Public Administration* 94 (2): 433–51. doi: 10.1111/padm.12224.
- Van der Voet, Joris, and Bram Steijn. 2020. Team innovation through collaboration: how visionary leadership spurs innovation via team cohesion. *Public Management Review*. doi: 10.1080/14719037.2020.1743344.
- van Dierendonck, Dirk. 2011. Servant Leadership: A Review and Synthesis. *Journal of Management* 37 (4): 1228–61. doi: 10.1177/0149206310380462.
- Van Wart, Montgomery. 2013. Administrative Leadership Theory: A Reassessment After 10 Years. *Public Administration* 91 (3): 521–43. doi: 10.1111/padm.12017.
- Viechtbauer, Wolfgang. 2010. Conducting meta-analyses in R with the metafor package. *Journal of Statistical Software* 36 (3): 1–48. <http://www.jstatsoft.org/v36/i03/>.
- Vogel, Dominik. 2018. Matching survey responses with anonymity in environments with privacy concerns: A Practical Guide. *International Journal of Public Sector Management* 31 (7): 742–54. doi: 10.1108/IJPSM-12-2017-0330.
- Vogel, Dominik, and Alexander Kroll. 2019. Agreeing to disagree? Explaining self-other disagreement on leadership behaviour. *Public Management Review* 21 (12): 1867–92. doi: 10.1080/14719037.2019.1577910.

- Vogel, Dominik, Artur Reuber, and Rick Vogel. 2020. Developing A Short Scale To Assess Public Leadership. *Public Administration* 98 (4): 958–73. doi: 10.1111/padm.12665.
- Vogel, Rick, and Doris Masal. 2015. Public Leadership: A review of the literature and framework for future research. *Public Management Review* 17 (8): 1165–89. doi: 10.1080/14719037.2014.895031.
- Wagenmakers, Eric-Jan, Tom Lodewyckx, Himanshu Kuriyal, and Raoul Grasman. 2010. Bayesian hypothesis testing for psychologists: A tutorial on the Savage–Dickey method. *Cognitive Psychology* 60 (3): 158–89. doi: 10.1016/j.cogpsych.2009.12.001.
- Wagenmakers, Eric-Jan, Jonathon Love, Maarten Marsman, Tahira Jamil, Alexander Ly, Josine Verhagen, Ravi Selker et al. 2018. Bayesian inference for psychology. Part II: Example applications with JASP. *Psychonomic Bulletin & Review* 25 (1): 58–76. doi: 10.3758/s13423-017-1323-7.
- Wasserstein, Ronald L., and Nicole A. Lazar. 2016. The ASA Statement on p -Values: Context, Process, and Purpose. *The American Statistician* 70 (2): 129–33. doi: 10.1080/00031305.2016.1154108.
- Wright, Bradley E., Shahidul Hassan, and Robert K. Christensen. 2017. Job Choice and Performance: Revisiting Core Assumptions about Public Service Motivation. *International Public Management Journal* 20 (1): 108–31. doi: 10.1080/10967494.2015.1088493.
- Wright, Bradley E., Donald P. Moynihan, and Sanjay K. Pandey. 2012. Pulling the Levers: Transformational Leadership, Public Service Motivation, and Mission Valence. *Public Administration Review* 72 (2): 206–15. doi: 10.1111/j.1540-6210.2011.02496.x.

APPENDIX A: OPERATIONALIZATIONS

A.1 Danish high schools

Construct	Operationalization
Transformational leadership (Podsakoff, MacKenzie, and Bommer 1996; Trottier, Van Wart, and Wang 2008; Wright, Moynihan, and Pandey 2012) ($\alpha = .92$)	As a leader, my principal provides a compelling vision of the organization's future ... articulates and generates enthusiasm for a shared vision and mission. ... facilitates acceptance of common goals for the school. ... say(s) things that make employees proud to be part of the organization.
Contingent rewards (Trottier, Van Wart, and Wang 2008; Hartog, van Muijen, and Koopman 1997) ($\alpha = .85$)	As a leader, my principal reward his employees' performance (e.g. through wage supplements) when they live up to expectations ... rewards his employees dependent on how well they perform their jobs
Sanctioning leadership (Trottier, Van Wart, and Wang 2008; Hartog, van Muijen, and Koopman 1997) ($\alpha = .55$)	As a leader, my principal focuses on irregularities, mistakes, exceptions and deviations from what is expected of me. ... dismisses teachers, if they do not perform satisfactorily over an extended period.

A.2. German state agencies

Construct	Operationalization
Transformational leadership (Wright, Moynihan, and Pandey 2012) ($\alpha = .88$)	My supervisor clearly articulates his/her vision of the future ... leads by setting a good example ... challenges me to think about old problems in new ways ... says things that make employees proud to be part of the organization ... has a clear sense of where our organization should be in five years

A.3 Danish schools, daycare centers, and tax offices

Construct	Operationalization
Transformational leadership (Jensen et al. 2019) ($\alpha = .89$)	<p>My leader ...</p> <p>... concretizes a clear vision for the [ORGANIZATION'S] future.</p> <p>... seeks to make employees accept common goals for the [ORGANIZATION].</p> <p>...strives to get the [ORGANIZATION'S] employees to work together in the direction of the vision.</p> <p>... strives to clarify for the employees how they can contribute to achieving the [ORGANIZATION'S] goals.</p>
Transactional leadership: Nonpecuniary rewards (Jensen et al. 2019) ($\alpha = .94$)	<p>My leader ...</p> <p>... gives individual employees positive feedback when they perform well.</p> <p>... actively shows his or her appreciation of employees who do their jobs better than expected.</p> <p>... personally compliments employees when they do outstanding work.</p>
Transactional leadership: Pecuniary rewards (Jensen et al. 2019) ($\alpha = .88$)	<p>My leader ...</p> <p>... rewards the employees' performance when they live up to his or her requirements</p> <p>... rewards the employees' dependent on how well they perform their jobs.</p> <p>... points out what employees will receive if they do what is required.</p>
Transactional leadership: Sanctioning leadership (Jensen et al. 2019) ($\alpha = .88$)	<p>My leader ...</p> <p>... gives negative consequences to the employees if they perform worse than their colleagues.</p> <p>... makes sure that it has consequences for the employees if they do not consistently perform as required.</p> <p>... gives negative consequences to employees if they do not perform as he or she requires.</p>

A.4 Dutch local welfare teams

Construct	Operationalization
Transformational leadership (Podsakoff, MacKenzie, and Bommer 1996; Trottier, Van Wart, and Wang 2008; Wright, Moynihan, and Pandey 2012) ($\alpha = .91$)	As a leader, my principal provides a compelling vision of the organization's future ... articulates and generates enthusiasm for a shared vision and mission. ... facilitates acceptance of common goals for the school. ... say(s) things that make employees proud to be part of the organization.

A.5 German city district offices

Construct	Operationalization
Public leadership (Tummers and Knies 2016) ($\alpha = .94$)	My leader encourages me and my colleagues to explain our actions to various stakeholders. ... encourages us to inform stakeholders of our way of working. ... provides us with the possibility to explain our behavior to stakeholders. ... emphasizes that it is important that we answer questions from clients ... strives to ensure that we openly and honestly share the actions of our organizational unit with others. ... encourages us to explain to stakeholders why certain decisions were taken. ... emphasizes to me and my colleagues that it is important to follow the law. ... gives me and my colleagues the means to properly follow governmental rules and regulations. ... emphasizes that my colleagues and I should carry out government policies properly. ... ensures that we accurately follow the rules and procedures.

-
- ... encourages me and my colleagues to support political decisions, even when other stakeholders confront us with it.
 - ... encourages me and my colleagues not to jeopardize the relationship with political heads, even if that entails risks.
 - ... encourages me and my colleagues to implement political decisions, even if that means undertaking additional responsibilities.
 - ... encourages me and my colleagues to defend political choices, even if we see shortcomings.
 - ... encourages me and my colleagues to support political decisions, even when we see downsides.
 - ... encourages me and my colleagues to maintain many contacts with other organizations.
 - ... encourages me and my colleagues to invest substantial energy in the development of new contacts.
 - ... motivates me and my colleagues to regularly work together with people from our networks.
 - ... motivates me and my colleagues to develop many contacts with people outside our own department.
 - ... encourages me and my colleagues to introduce others to contacts of our own networks.

Authentic leadership
(Neider and Schriesheim
2011) ($\alpha = .93$)

- My leader ...
- ... clearly states what he/she means.
 - ... shows consistency between his/her beliefs and actions.
 - ... asks for ideas that challenge his/her core beliefs.
 - ... describes accurately the way that others view his/her abilities.
 - ... uses his/her core beliefs to make decisions.
 - ... carefully listens to alternative perspectives before reaching a conclusion.
 - ... shows that he/she understands his/her strengths and weaknesses.
 - ... openly shares information with others.
 - ... resists pressures on him/her to do things contrary to his/her beliefs.
 - ... objectively analyzes relevant data before making a decision.
 - ... is clearly aware of the impact he/she has on others.
 - ... expresses his/her ideas and thoughts clearly to others.
 - ... is guided in his/her actions by internal moral standards.
-

... encourages others to voice opposing points of view.

A.6 Dutch unemployment insurance agency

Construct	Operationalization
Servant leadership (van Dierendonck and Nuijten 2011) ($\alpha = .95$)	<p>My team manager ...</p> <ul style="list-style-type: none">... gives me the information I need to do my work well.... encourages me to use my talents.... helps me further develop myself.... encourages his/her staff to come up with new ideas.... gives me the authority to take decisions which make work easier for me.... enables me to solve problems myself instead of just telling me what to do.... offers me abundant opportunities to learn new skills.... keeps him-/herself in the background and gives credit to others.... is not chasing recognitions or rewards for the change the things he/she does for others.... appears to enjoy his/her colleagues' success more than his/her own.... holds me responsible for the work I carry out. <p>I am held accountable for my performance by my team manager.</p> <ul style="list-style-type: none">... holds me and my colleagues responsible for the way we handle a job.... keeps criticizing people for the mistakes they have made in their work. (R)... maintains a hard attitude towards people who have offended him/her at work. (R)... finds it difficult to forget things that went wrong in the past. (R)... takes risks even when he/she is not certain of the support from his/her own manager.... takes risks and does what needs to be done in his/her view.... is open about his/her limitations and weaknesses.

... is often touched by the things he/she sees happening around him/her.

... is prepared to express his/her feelings even if this might have undesirable consequences.

... shows his/her true feelings to his/her staff.

... learns from criticism.

... tries to learn from criticism he/she gets from his/her superior.

... admits his/her mistakes to his/her superior.

... learns from the different views and opinions of others.

If people express criticism, my team manager tries to learn from it.

emphasizes the importance of focusing on the good of the whole.

... has a long-term vision.

... emphasizes the societal responsibility of our work.

Public leadership: rule-following (Tummers and Knies 2016) ($\alpha = .97$)

My team manager ...

... emphasizes to me and my colleagues that it is important to follow the law.

... gives me and my colleagues the means to properly follow ... governmental rules and regulations.

... emphasizes that my colleagues and I should carry out government ... policies properly.

... ensures that we accurately follow the rules and procedures.

APPENDIX B: DETAILED RESULTS

B.1 Danish high schools

<i>Predictors</i>	Transformational leadership			Rewarding leadership			Sanctioning Leadership		
	<i>Estimates</i>	<i>std. Error</i>	<i>p</i>	<i>Estimates</i>	<i>std. Error</i>	<i>p</i>	<i>Estimates</i>	<i>std. Error</i>	<i>p</i>
Intercept	2.912	0.064	<0.001	2.465	0.052	<0.001	2.341	0.056	<0.001
Leader response	0.142	0.080	0.076	0.005	0.065	0.933	-0.041	0.069	0.552
Random Effects									
σ^2	0.83			0.96			0.73		
τ_{00}	0.19			0.11			0.14		
n (schools)	161			161			161		
n (teachers)	4255			4255			4255		
log-Likelihood	-5780.767			-6050.284			-5509.765		

B.2 German state agencies

<i>Predictors</i>	Transformational leadership		
	<i>Estimates</i>	<i>std. Error</i>	<i>p</i>
Intercept	2.737	0.094	<0.001
Leader response	0.238	0.135	0.080
Random Effects			
σ^2	0.79		
τ_{00}	0.27		
n (teams)	118		
n (employees)	471		
log-Likelihood	-661.178		

B.3 Danish schools, daycare centers, and tax offices

<i>Predictors</i>	Transformational leadership			Nonpecuniary rewards			Pecuniary rewards			Sanctioning Leadership		
	<i>Estimates</i>	<i>std. Error</i>	<i>p</i>	<i>Estimates</i>	<i>std. Error</i>	<i>p</i>	<i>Estimates</i>	<i>std. Error</i>	<i>p</i>	<i>Estimates</i>	<i>std. Error</i>	<i>p</i>
Intercept	4.073	0.151	<0.001	3.732	0.179	<0.001	2.416	0.151	<0.001	2.756	0.113	<0.001
Leader response	-0.242	0.153	0.116	-0.152	0.181	0.402	0.125	0.152	0.415	0.129	0.114	0.257
Random Effects												
σ^2	0.59			0.92			0.75			0.63		
τ_{00}	0.17			0.23			0.15			0.07		
n (teams)	378			378			378			378		
n (employees)	6365			6365			6365			6365		
log-Likelihood	-7660.480			-9052.146			-8376.105			-7749.467		

B.4 Dutch local welfare teams

<i>Predictors</i>	Transformational leadership		
	<i>Estimates</i>	<i>std. Error</i>	<i>p</i>
Intercept	3.705	0.145	<0.001
Leader response	-0.047	0.150	0.754
Random Effects			
σ^2	0.57		
τ_{00}	0.15		
n (teams)	182		
n (employees)	1358		
log-Likelihood	-1638.784		

B.5 German city district offices

<i>Predictors</i>	Public Leadership			Authentic Leadership		
	<i>Estimates</i>	<i>std. Error</i>	<i>p</i>	<i>Estimates</i>	<i>std. Error</i>	<i>p</i>
Intercept	3.312	0.027	<0.001	3.354	0.033	<0.001
Leader response	0.094	0.040	0.020	0.085	0.048	0.079
Random Effects						
σ^2	0.43			0.53		
τ_{00}	0.13			0.20		
n (teams)	698			698		
n (employees)	2274			2274		
log-Likelihood	-2477.101			-2763.703		

B.6 Dutch unemployment insurance agency

<i>Predictors</i>	Servant leadership			Rule-following leadership		
	<i>Estimates</i>	<i>std. Error</i>	<i>p</i>	<i>Estimates</i>	<i>std. Error</i>	<i>p</i>
Intercept	5.092	0.042	<0.001	4.910	0.067	<0.001
Leader response	-0.014	0.062	0.824	0.053	0.100	0.601
Random Effects						
σ^2	0.62			1.63		
τ_{00}	0.13			0.34		
n (teams)	277			277		
n (employees)	1602			1602		
log-Likelihood	-2002.539			-2768.457		