

Backhaus, Leonie; Vogel, Rick

Article — Published Version

Leadership in the public sector: A meta-analysis of styles, outcomes, contexts, and methods

Public Administration Review

Provided in Cooperation with:

John Wiley & Sons

Suggested Citation: Backhaus, Leonie; Vogel, Rick (2022) : Leadership in the public sector: A meta-analysis of styles, outcomes, contexts, and methods, Public Administration Review, ISSN 1540-6210, Wiley Subscription Services, Inc., Hoboken, USA, Vol. 82, Iss. 6, pp. 986-1003, <https://doi.org/10.1111/puar.13516>

This Version is available at:

<https://hdl.handle.net/10419/266754>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.



<http://creativecommons.org/licenses/by-nc/4.0/>

Leadership in the public sector: A meta-analysis of styles, outcomes, contexts, and methods

Leonie Backhaus  | Rick Vogel 

Faculty of Business, Economics and Social Sciences, Department of Socioeconomics, Chair of Public Management, Universität Hamburg, Hamburg, Germany

Correspondence

Leonie Backhaus, Faculty of Business, Economics and Social Sciences, Department of Socioeconomics, Chair of Public Management, Universität Hamburg, Von-Melle-Park 9, 20146 Hamburg, Germany.
Email: leonie.backhaus@uni-hamburg.de

Abstract

Leadership has attracted growing attention among scholars and practitioners in public administration. With the rising availability of study results, however, it becomes increasingly difficult to keep track under which conditions leadership does or does not make a difference in the public sector. This study provides a meta-analysis of administrative leadership and various correlates that research has theorized as outcomes of leadership. The results of a multi-level random-effects meta-analysis based on 486 effect sizes from 151 studies show that correlations are stronger for the achievement of beneficial than for the prevention of detrimental outcomes, as well as for group- and organization-related than for employee-related outcomes. Moderation analyses reveal that leadership style, administrative tradition, administrative subfield, and methodological factors explain heterogeneity in effect sizes.

Evidence for Practice

- Leadership in the public sector is positively and largely consistently related to beneficial outcomes, such as performance, and negatively associated with detrimental outcomes, such as turnover.
- Within the range of the analyzed leadership styles, it is not particularly important *how* public leaders lead — as long as they *do* lead. Laissez-faire leadership is confirmed to be a largely ineffective style of leadership. Leaders should be similarly cautious with the exercise of controlling forms of leadership.
- The broad range of leadership styles helps HR practitioners to adjust leadership trainings to the available talents and given settings. Public leaders may feel encouraged that there are many ways to become effective leaders.
- Administrative tradition matters, as the leadership-outcome relationship is less strong in some continental European as well as East Asian traditions than in the Anglo-American tradition. For example, this applies to transformational leadership.

INTRODUCTION

Leadership is among the core concepts in public administration (PA) scholarship (Chapman et al., 2016; Crosby & Bryson, 2018; Ospina, 2017; Van Wart, 2003, 2013a). Two decades ago, research on leadership in the public sector was at a nascent but still limited stage, severely lagging behind general leadership studies. Van Wart (2003), in his review of the then available literature, arrived at the conclusion that “the needs are great and the research opportunities are

manifold” (p. 225). Since then, PA scholars have frequently echoed this call and seized the opportunities for leadership research (Chapman et al., 2016; Van Wart, 2013a; Vogel & Masal, 2015). The number of available studies has grown considerably, forming a flourishing stream in PA scholarship.

The growing availability and diversity of studies, however, does not make it easier to answer one of the core questions of leadership: “To what degree does leadership make a difference?” (Van Wart, 2003, p. 221). Although scholars and practitioners would largely agree that leadership is an important

This is an open access article under the terms of the [Creative Commons Attribution-NonCommercial](https://creativecommons.org/licenses/by-nc/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.

© 2022 The Authors. *Public Administration Review* published by Wiley Periodicals LLC on behalf of American Society for Public Administration.

phenomenon that can make a difference in public organizations, there is little clarity beyond this general consent. The polyphony begins with different conceptualizations of leadership characteristics and behaviors. A majority of scholars has focused on transformational leadership (Chapman et al., 2016), but many more leadership styles have recently been examined. Some scholars have voiced concerns about the adoption of generic leadership styles and have begun to develop leadership constructs that account for the specificities of public leadership (Fernandez et al., 2010; Tummers & Knies, 2016; Vogel et al., 2020). There is thus a plethora of leadership styles, and it is unclear if there are differences in the difference they potentially make.

Conceptual diversity does not only characterize leadership as an independent variable, but also the outcomes of leadership, that is, the dependent variables in the leadership equation. Consistent with an understanding of leadership as a process of influencing others (Yukl & Gardner, 2020), many scholars have examined how leadership influences the attitudes, motivations, and behaviors of followers, such as job satisfaction (e.g., Park & Rainey, 2008), work engagement (e.g., Ancarani et al., 2020), or individual performance (e.g., Hassan, Park, & Raadschelders, 2019). However, the outcomes differ strongly in kind and may relate, for instance, to employees (such as work engagement; Tummers & Knies, 2013), groups (such as team climate; Kroll & Vogel, 2014), or the organization (such as organizational performance; Jakobsen & Andersen, 2015). Furthermore, leadership may not only be effective in bringing about beneficial outcomes but also in preventing detrimental outcomes, such as turnover intentions (e.g., Park & Rainey, 2008). It is unclear, however, if leadership makes the same difference across outcomes of different kinds and valence.

With the growth of leadership studies in PA, more insights from various subfields of the public sector and from different national sites have become available. While this diversity is both needed and welcomed, it also raises the question whether leadership is effective in different contexts to similar extents (e.g., Slyke et al., 2006). For example, leadership has been studied in subfields of the public sector as diverse as the federal government (e.g., Caillier, 2014), education (e.g., Bauwens et al., 2019), and the military (e.g., Hattke et al., 2018). The emergence and effectiveness of leadership may be contingent on the distinct professional cultures in each of these subfields. Leadership may also co-vary with national and administrative traditions and cultures (House et al., 2004; Painter & Peters, 2010). Administrative systems are likely to be characterized by different configurations of “substitutes for leadership” (Kerr & Jermier, 1978), which may enhance or neutralize the effects of leadership. In short, the extent to which insights from leadership studies can be generalized and transferred across different administrative subfields and systems remains an open issue. It is therefore fair to say that scholarship has not yet fully accounted for the context in which leadership occurs (e.g., Ospina, 2017; Van Wart, 2013a; Vogel & Masal, 2015).

Considerable variation occurs not only in the *what* and *where* of studying leadership in the public sector, but also in the *how*. Leadership scholars can choose from a large pool of research instruments and designs when they decide how to gather data and how to arrange the analysis. Among the decisions with which researchers frequently struggle is whom to ask for leadership ratings, which data sources to combine, and when to measure the variables. Such decisions may have important consequences for the results that will be obtained and for the claims that can be made. In the broader PA community, there is increasing awareness of these issues, most evidently reflected in discussions about common method bias (George & Pandey, 2017; Jakobsen & Jensen, 2015; Meier & O’Toole, 2013). How methodological factors interfere with the association between leadership and outcomes, however, has not yet been assessed on a broad scale.

To conclude, readers of PA journals face a growing body of leadership research but are increasingly left puzzled with what they can learn from single studies and pass on to practitioners, given the variety of styles, outcomes, contexts, and methods in this scholarship. The present study aims to synthesize previous findings in an up-to-date and integrative review of quantitative leadership research in the public sector. The focus is on administrative leadership, that is, the leadership of non-elected leaders in public sector settings, rather than on political or community leadership (Van Wart, 2013a). *To what extent is administrative leadership related to outcomes of different kinds and valence? How do leadership styles as well as administrative traditions and subfields moderate these relationships? Which role do methodological factors, such as rating methods, data sources, and research designs, play in explaining variance in the leadership-outcome relationship?*

To address these questions, this article presents a meta-analysis of the relationship between administrative leadership and various correlates that previous scholarship has conceptualized as outcomes of leadership. The study builds on 151 studies ($N = 2,819,591$) published in PA journals since 2000 and conducts a multi-level random-effects meta-analysis with moderator analyses. The detailed findings improve the development of theory and research and inform administrative practice. Scholars learn about under-researched phenomena and important contingencies of leadership in the public sector, thus being provided with guidance in setting the agenda for future research and moving the field forward. Practitioners can conclude which leadership styles are worth developing and where to pay attention when transferring these styles from one context to another.

LEADERSHIP IN THE PUBLIC SECTOR: STYLES, OUTCOMES, CONTEXTS, AND METHODS

This section develops the conceptual framework for the meta-analysis along the four lines introduced above, each addressing an important question that needs to be

answered in leadership studies: what to study on both sides of the leadership equation (i.e., styles and outcomes), where to study it (i.e., contexts), and how to study it (i.e., methods). These questions arise for each individual researcher but also relate to broader discussions in the public leadership community and beyond. The framework balances this relevance with the practical requirements of a meta-analysis, as only those aspects can be examined that have been reported in primary studies and are available in sufficient numbers.

Leadership styles

Few scholars and practitioners would disagree that leadership can make a difference in public organizations. However, it is less clear what kind of leadership makes a difference in what outcomes, and how significant this difference is. Along with the considerable growth in publications over the past two decades, PA scholarship has examined an increasingly wide range of leadership concepts. This diversification particularly applies to leadership styles, that is, the “mid-range conceptualization of recognizable patterns of leader characteristics and behaviors” (Van Wart, 2013b, p. 531). Leadership styles are thus characteristics and behaviors of leaders which recur in consistent combinations. While the adoption and development of new leadership styles expand the field’s variety and richness, it may also facilitate a process that has been referred to as a “Balkanization of the field with innumerable aspects of leadership, each with its own special and often conflicting terms” (Van Wart, 2013b, p. 537). As a result, scholars increasingly face the issue of how the various leadership styles relate to and are distinct from each other in terms of both conceptualization and measurement. Practitioners are left in doubt about which of the many styles to prioritize when it comes to developing and promoting leadership within given constraints of time and resources.

The range of leadership styles discussed in PA scholarship is indeed impressive and has recently further expanded, for instance, entrepreneurial leadership (e.g., Miao et al., 2018), ethical leadership (e.g., Hassan et al., 2014), servant leadership (e.g., Schwarz et al., 2016), shared leadership (e.g., Umans et al., 2018), and paradoxical leadership (e.g., Backhaus et al., 2021). These styles are very likely to have aspects in common, but also vary in the source, object, and result of leadership (Ospina, 2017). The most widely discussed style of leadership, both in PA scholarship and beyond, is transformational leadership (e.g., Jacobsen & Staniok, 2020; Sun & Wang, 2017; Wright & Pandey, 2010), figuring as an individual substream in public leadership research (Vogel & Masal, 2015). In this case, the leader is the source, employees are the “object,” and heightened levels of motivation and empowerment are the result of leadership (Ospina, 2017). Among the reasons why transformational leadership receives both scholarly and practical

attention is that studies have repeatedly proven that it actually delivers these results (e.g., Ancarani et al., 2020; Muterera et al., 2018). It is unclear, however, if other leadership styles deliver under the same conditions to a similar extent, as only a few studies investigate multiple styles simultaneously.

Leadership outcomes

The extent to which leadership styles attract scholarly and practical attention depends largely on their effectiveness in bringing about desired outcomes in public organizations. Therefore, many if not most studies in the field relate leadership styles to various outcome variables as cornerstones of effectiveness (Vogel & Masal, 2015). A prime example is the association between leadership styles and job satisfaction which is among the best-researched attitudes of public employees (Cantarelli et al., 2016). The common saying that people quit their bosses, rather than their jobs, underscores the pivotal role of leaders in employees’ job satisfaction. The PA scholarship is largely consistent in establishing statistically significant relationships between leadership and job satisfaction of followers, but the strength of this association may vary considerably across leadership styles. For example, Muterera et al. (2018) focus on transformational leadership and find an exceptionally strong correlation with job satisfaction ($r = .83$), while Park and Rainey (2008) examine transactional leadership and report only a weak correlation ($r = .30$). These anecdotal pieces from the literature illustrate that different leadership styles may relate to the same outcomes in very different strengths. They also question whether single studies allow for any general conclusions on the effectiveness of leadership styles.

The effectiveness of the same leadership style may also vary considerably across different kinds of outcomes. Consistent with an understanding of leadership as a process of influencing others (Yukl & Gardner, 2020), most studies in the field focus on the implications of leadership on attitudes, motivations, and behaviors of employees (e.g., Hassan, Park, & Raadschelders, 2019; Miao et al., 2018; Wright et al., 2012). Leadership outcomes may also relate to jobs, such as perceptions of job autonomy (Tummers et al., 2018), or to managerial practices, such as the use of performance information (Masal & Vogel, 2016). However, scholars and practitioners are often even more interested in the consequences of leadership at collective levels, such as group climate (Hassan et al., 2014), organizational performance (Jacobsen & Andersen, 2015), or societal impact (Luu, 2018b). These are less proximal and more distal outcomes of leadership, and it is unclear whether leadership yields the same effects across all these outcomes, and how leadership styles differ in this regard.

Leadership outcomes also vary in terms of desirability. For example, outstanding performance of followers has arguably a positive valence as a leadership outcome (Hassan, Park, & Raadschelders, 2019), while ethical misbehavior is

clearly among the negative workplace behaviors that should be prevented (Hassan et al., 2014). Scholars and practitioners usually consider leadership effective when it facilitates beneficial outcomes but inhibits detrimental outcomes, but the underlying social-psychological mechanisms through which leadership yields these effects may be very different. Hence, the implications of leadership in general and the various leadership styles in particular are also likely to differ depending on the valence of outcome variables (i.e., beneficial or detrimental).

Leadership context

The previous sections suggest that there is heterogeneity both in the association of different leadership styles with the same outcome and in the relationship of the same style with different outcomes. Moreover, and probably more puzzling, leadership studies are also remarkably heterogeneous in how the same leadership style is associated with the same outcome. For example, while Muterera et al. (2018) report a very strong correlation between transformational leadership and job satisfaction ($r = .83$), the same relationship is much weaker ($r = .30$) in the study by Masal and Vogel (2016). In explaining such differences, the context in which leadership occurs is likely to be an important yet often neglected factor. Scholars have repeatedly stressed the need to understand leadership as a context-dependent phenomenon (Ospina, 2017; Vogel & Masal, 2015; Wright & Pandey, 2010). Context is a blind spot not only in public leadership studies, but also in PA scholarship more generally (Meier et al., 2017; O'Toole & Meier, 2014). As a consequence, insights into the generalizability and transferability of research findings and their practical implications across different contexts remain limited.

While many contextual variables may have an influence on the leadership-outcome link, few of them are both of broader relevance and available from the reporting in leadership studies. Administrative subfields and traditions are two important exceptions worth considering here: First, leadership contexts differ profoundly across *administrative subfields*. While Muterera et al. (2018) conducted their study on county governments, Masal and Vogel (2016) examined leadership in police force. These fields differ strongly in, for example, task characteristics, professional cultures, and organizational structures. The police is a prime example of “command-and-control” organizations in which the effectiveness of transformational leadership may be limited, as law enforcement operations may require clear instructions rather than charismatic visions. In contrast, the provision of public service at the county level may set a better stage for sense-giving by transformational leaders. Indeed, substitutes for leadership theory suggests that situational factors may enhance, neutralize, or entirely substitute for leadership (Kerr & Jermier, 1978). Scholarship in PA is thus well advised to account for the context of administrative subfields when

leadership is of interest (Slyke et al., 2006). This applies all the more as scholars studying administrative subfields often form their own subcommunities, with often undiscovered potential to engage in exchange with each other. Insights into the commonalities and differences of leadership across subfields help to assess this potential.

Second, leadership may also differ across *administrative traditions*. With the growth of publications over the past decades, more leadership studies from a broader range of countries have become available (Van Wart et al., 2015). However, the extent to which insights from these studies can be transferred from one national site to another remains unclear. Administrative systems have evolved over long periods of time within broader administrative traditions (Kuhlmann & Wollmann, 2019; Painter & Peters, 2010), and the idiosyncrasies of these traditions may matter for the emergence and effectiveness of leadership. For example, Muterera et al. (2018) collected data in the US, whereas Masal and Vogel (2016) conducted their study in Germany. The US is a prime example of the Anglo-American administrative tradition in which management and policy education plays a major role and emphasis is put on pragmatism, flexibility, and managerial autonomy. In contrast, Germany represents the continental European tradition with primacy on the legal profession and more emphasis on legality, hierarchy, and neutrality (Kuhlmann & Wollmann, 2019; Painter & Peters, 2010). Scholarship in PA has repeatedly acknowledged such differences as important antecedents of administrative structures and behavior (Mohr et al., 2021; Raudla et al., 2021; Van der Wal et al., 2021) and therefore considered administrative traditions in other meta-analyses (George et al., 2021).

Research methods

Scholarship also varies considerably in how leadership in the public sector is studied. While surveys are still the overall method of choice in a vast majority of studies, there are many questions left in the detailed configuration and implementation of questionnaires. The first issue is whom to ask about leadership. For example, Muterera et al. (2018) have asked both leaders and followers for their ratings of the leader's transformational leadership. They find an exceptionally high correlation with job satisfaction when leadership is rated by followers ($r = .83$), but this correlation drops considerably when leaders rate their own behaviors ($r = .26$). Public management scholars have elaborated on the substantial differences between self- and other-ratings of leadership, providing vast support for leaders' tendency to overestimate their behaviors as compared to employee-perceived leadership (Jacobsen & Andersen, 2015; Vogel & Kroll, 2019). Therefore, the type of leadership rating (i.e., self- or other-ratings) should be considered as a methodological factor that may explain heterogeneity in effects across leadership studies.

A second and related issue is the choice and combination of data sources. Researchers may ask the same respondents for both the leadership and the outcome variable, or they may ask different informants and combine their responses afterward. The former case gives rise to severe concerns of common source bias, as self-reported measures from the same source may inflate results (George & Pandey, 2017; Jakobsen & Jensen, 2015; Meier & O'Toole, 2013). Accordingly, leadership scholars may have strong reasons for asking different people, or not to ask people at all when measuring the outcome variable. Perceptual measures of leadership, whether they are self- or other-rated, may be combined with objective measures of the outcome, such as actual turnover (Tavares et al., 2021). In any case, the use of multiple data sources is likely to make a difference compared to the standard case of a single source.

Finally, researchers have the choice between different research designs into which the data collection is embedded. Most leadership studies have applied cross-sectional designs, with the independent and dependent variables being measured at the same point in time. To mitigate concerns of common method bias associated with these designs (George & Pandey, 2017; Meier & O'Toole, 2013), it has been suggested to measure independent and dependent variables at different points in time (Podsakoff et al., 2003). In case of such a temporal separation, a time-lagged or longitudinal design is in place. Leadership scholars may also renounce survey research and join the recent plethora of experimental studies in PA scholarship (e.g., Bellé, 2014; Jacobsen et al., 2021). In short, heterogeneity in effects across leadership studies may have origins in the applied rating methods, data sources, and research designs, but how significant these differences are have not yet been assessed in a broad range of studies.

Figure 1 presents the conceptual framework for the meta-analysis resulting from the above lines of reasoning. Technically, the analysis focuses on main and moderation effects. Following previous meta-analyses in leadership studies (e.g., Schyns & Schilling, 2013), the framework makes a distinction between beneficial and detrimental outcomes, although this is a simplification of what is likely to be a continuum in

reality. Where the number of leadership studies is already sufficient, post-hoc analyses can zoom in on the association of leadership with specific outcomes. If and how leadership in the public sector translates into outcomes is contingent on several factors, which have been introduced above and enter the conceptual framework as moderators (i.e., leadership styles, leadership outcomes, leadership contexts, research methods). Of course, many more factors may moderate the relationship between leadership and outcomes. The focus and resolution of the present meta-analysis are adjusted to the field's state-of-the-art, addressing those moderators that are both relevant to ongoing discussions in the community and available from a critical mass of primary studies.

METHODS

Meta-analyses have recently gained momentum in PA scholarship (e.g., George et al., 2021; Homberg et al., 2015; Park, 2020), as they allow for a systematic aggregation of effects across large numbers of samples and, in turn, for an inspection of contextual contingencies of these effects. The reader should be aware, however, that this meta-analysis is based on a correlational technique. If primary studies do not use causal research designs, meta-analyses cannot establish causal effects. Referring to correlates as "leadership outcomes," and concluding from such associations on leadership effectiveness, is a conceptual rather than a methodological choice. This choice resonates with a large body of scholarship that has provided solid theoretical foundations and vast empirical evidence for such leadership effects (Yukl & Gardner, 2020). It is also consistent with the recommendation of Ringquist (2013) to use correlations as effect sizes for meta-analyses in the field of PA and with the vocabulary of other meta-analyses (e.g., George et al., 2021).

Data

The primary studies for the meta-analysis were selected from journals in the "Public Administration" subject

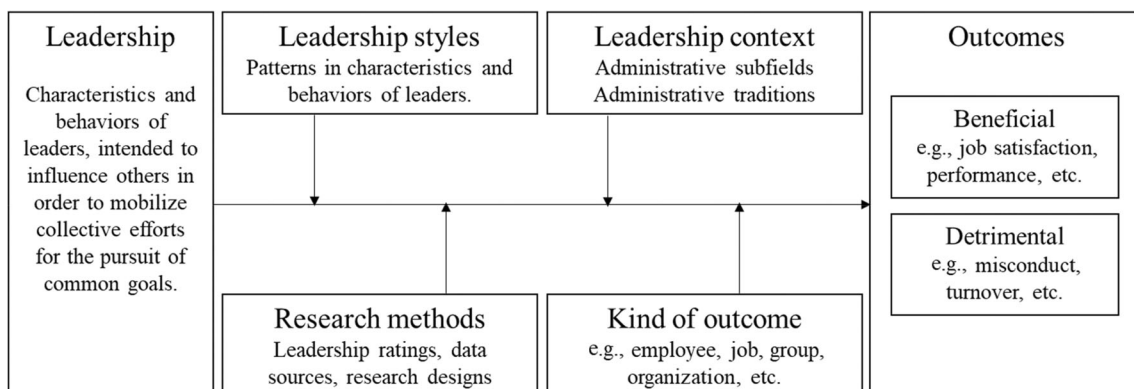


FIGURE 1 Framework for meta-analysis

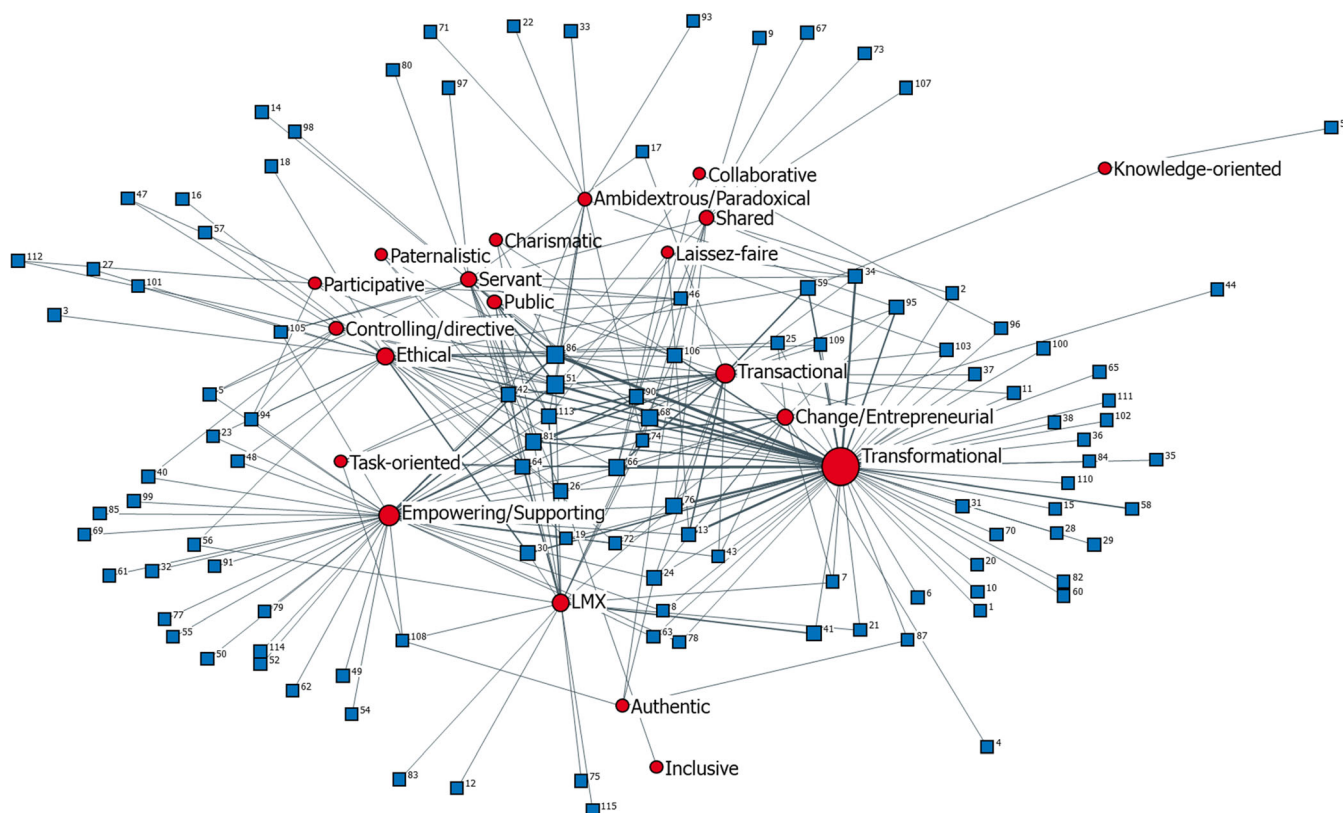
category in the Social Science Citation Index (online Appendix A). This database mainly covers scholarly, peer-reviewed journals, including the field's top journals, which should ensure a high quality of the data that feed into the meta-analysis. The search query covered all articles from 2000 onwards in order to maintain recency. Within the body of more than 25,000 journal articles published in English language in the PA category since 2000, 2377 (around 9%) articles referred to "leadership" in the title, keywords, and/or abstract. The authors screened these articles and selected those including quantitative analyses of leadership in conjunction with leadership outcomes. A requirement for inclusion was the reporting of correlation coefficients for the association between leadership and outcome variables. The same applies to sample sizes. If correlations and/or sample sizes were not reported, the authors of the original studies were contacted and asked for the missing information. The final database consists of 486 effect sizes ($N = 2,819,591$) from 151 studies using 133 samples (see online Appendix B for all included studies). A flow chart illustrating the systematic literature review process is presented in online Appendix C.

Coding

This section describes the coding procedure in which the variables resulting from the conceptual framework (Figure 1) were extracted from the primary studies. Online Appendices D–E provide supplementary information on the applied coding categories. By presenting the data table, online Appendix N makes the results of the coding procedure for all included studies and effects available.

Leadership styles

The coding for leadership styles resulted in 20 categories (i.e., styles), for which well-established labels were adopted (Figure 2). Some styles merged into a broader category, as these styles differed in their original labels but showed large overlaps or strong similarities in their content. An example is "empowering leadership" (Hassan, DeHart-Davis, & Jiang, 2019) and "supportive leadership" (Hattke et al., 2018), which were merged into a joint category (i.e., "empowering/supportive"). Although leader-member exchange refers to the relationship between leaders and followers rather than to the characteristics and behaviors of leaders, it was included in



the coding list, as there is growing interest in relational approaches to leadership in the public sector (Ospina, 2017). Some studies measure leadership with broad instruments covering multiple dimensions of leadership behaviors (e.g., Elkomy et al., 2020). These studies were considered in the meta-analysis of main effects but excluded from the moderation analysis because the applied measurements and corresponding conceptualizations do not conform to the definition of leadership styles (i.e., specific patterns of leadership characteristics and behaviors).

Leadership outcomes

In total, 115 different outcomes were identified and coded in terms of two broader characteristics: First, the variables were coded as either *beneficial* or *detrimental outcomes* to inform the meta-analysis of the main effects (Figure 1). This coding was based on an assessment of the outcomes' valence (i.e., positive or negative), which was in most cases apparent from the label or description of the variables. However, in a few cases, outcomes were excluded from further analyses because they have no clear "sign" in terms of valence, such as, for example, enforcement styles of street-level bureaucrats (Klijn et al., 2020). Second, the moderation analysis required coding for the *kind of outcome*, with four categories: employee-related (e.g., individual performance), job-related (e.g., job autonomy), management-related (e.g., performance management), and group/organization-related (e.g., organizational performance). Grouping outcomes into broader categories is a common practice in meta-analyses both in PA (George et al., 2021) and beyond (Schyns & Schilling, 2013), particularly when the primary interest is not in a specific outcome but in the independent variable (which is the case here). Few studies have also examined leader-related outcomes (e.g., trust in leader; Asencio, 2016) and community-related outcomes (e.g., citizen value co-creation; Luu, 2018b), but the number of effect sizes for these kinds of outcomes was still too small to warrant meta-analysis. The assignment of outcomes to the coding categories is presented in online Appendix D.

Leadership context

Regarding the *administrative subfield*, the studies clustered in five fields: core administration (e.g., federal, state or local government, public agencies), education (e.g., public schools, universities), health (e.g., public hospitals, public health care settings), law enforcement/military (e.g., police, military), and social services (e.g., publicly employed social workers, social housing, welfare professionals). In some cases, data was collected across several subfields; these studies were included in the meta-analysis of main effects but excluded from the

moderation analysis. To code the *administrative tradition*, the authors applied the frequently used taxonomy by Painter and Peters (Painter & Peters, 2010) and assigned the countries in which the studies were conducted to one of the following traditions: Anglo-American, East Asian, Germanic, Islamic, Latin American, Napoleonic, Postcolonial South Asia and Africa, Scandinavian, and Soviet. A complete list of countries that were assigned to these traditions is provided in online Appendix E.

Research methods

The methods applied in leadership research in the public sector are reflected in three coding dimensions. For all these dimensions, it should be noted that the moderators were coded on the level of single effects, rather than on the level of studies, because the same study may include multiple effects building on different methods. The first dimension is the rating method of the leadership measure (*leadership rating*), which was binary coded into self-ratings (by leaders) or other-ratings (by others, mostly followers). Second, it was coded for the *data source*, indicating as to whether data for the independent and dependent variable was gathered from the same source (single-source) or from different sources (multi-source). The latter case occurs when leadership and outcomes are rated by different respondents or built on subjective and objective measures, respectively. Third, the *research design* was coded into cross-sectional or time-lagged designs, with the latter having a temporal separation between the measurement of leadership and outcomes. This coding scheme does not include "truly" longitudinal studies because they were too few in number to warrant a coding category on their own. In case longitudinal effects of leadership were available from repeated observations of the same respondents, they were assigned to the time-lagged category. Further, it could not be coded for experimental studies, although this would have been desirable. While there is a small but growing number of experimental studies on leadership in the public sector, most of these studies consider leadership as a dependent variable. This is evident in cases of field experiments on leadership trainings, where trainings are the intervention and leadership, and changes thereof, is the outcome (e.g., Jacobsen et al., 2021). In contrast, our conceptual framework (Figure 1) considers leadership as an independent variable. Only a few experimental studies resonate with this framework by manipulating leadership behaviors, for instance, by presenting scenarios of different leaders (for a rare exception, see Bellé, 2014). Even if these cases were sufficient in number, the reporting standards of experimental studies commonly do not require authors to calculate correlations between the manipulated leadership variable and outcomes.

Statistical procedure

The meta-analysis was conducted based on the recommendations of Ringquist (2013), Field and Gillet (2010), as well as Steel et al. (2021). The statistical procedure followed four steps: (1) extracting effect sizes, (2) conducting meta-analysis, (3) conducting moderator analyses, and (4) analyzing publication bias.

Extracting effect sizes

A meta-analysis calculates an overall population effect size based on single effect sizes retrieved from primary studies and weighted by the sample sizes. Please note that “effect size” is a common term in meta-analyses, even if the measures are acquired from studies with non-causal research designs. Indeed, using Pearson correlation coefficients as effect sizes offers the best comparability and applicability (e.g., Field & Gillett, 2010; George et al., 2021; Homberg et al., 2015), as it is common practice to report them in quantitative analyses. In the case of dimensional measures of leadership styles, some authors do not report correlations for the aggregate construct but only for the subdimensions. Consistent with conventions (Geyskens et al., 2008), these correlations were aggregated into a single correlation coefficient by calculating the mean across the partial correlations. The transformation of all Pearson product-moment correlation coefficients into a Fisher’s z -metric ($z = \frac{1}{2} \ln\left(\frac{1+r}{1-r}\right)$) corrects for the askew sampling distribution of Pearson correlation coefficients and ensures that the assumption of an approximately normal sampling distribution of the transformed values is justified (Ringquist, 2013). Although all calculations were based on Fisher’s z -metric, the meta-analytic results are presented in the form of retransformed Pearson product-moment correlations ($r = \frac{\exp(2z) - 1}{\exp(2z) + 1}$) to facilitate the interpretations of the results.

Conducting meta-analysis

To calculate the overall population effect size across all included articles, multi-level random effects meta-analytic procedures were performed with the *metafor* package in *R* (Viechtbauer, 2010). Applying multi-level meta-analysis accounts for dependencies resulting from the fact that most studies report several correlations based on the same sample (Konstantopoulos, 2011). Clustering the correlation coefficients within samples accounts for these dependencies. Accordingly, the model differentiates between three types of variance: sampling error variance (v) and two sources of true parameter variance (τ^2). These two sources of true parameter variance (i.e., the variance in the true underlying effect between samples) are: (1) the between-sample variance (σ_1^2), that is, the systematic differences between effects observed in different samples, and (2) the within-sample variance (σ_2^2), that is, the

systematic differences between multiple effects observed within the same sample. The mean effect size across samples and the variance components were estimated based on the restricted maximum likelihood estimator to assure that the true variance was not underestimated or affected by biases (Langan et al., 2019; Viechtbauer, 2010). Meta-analytic results based on mean effect sizes, 95% confidence intervals, and p -values were analyzed. Further, the Q -test for heterogeneity was used to identify whether effect sizes significantly vary across samples. A significant Q -test indicated the necessity to perform moderator analyses.

Conducting moderator analyses

To evaluate the moderating effects in the conceptual model, a meta-regression with dummy-coded moderator variables as predictors were conducted. A significant Q_M -statistic served as an indication of a significant moderator effect, as it indicates the presence of at least one significant difference in the mean effect size between a moderator category and the reference category. According to Ringquist (2013), this step is particularly important because in PA research the primary interest of meta-analysis is to investigate factors that account for heterogeneity in effect sizes.

Analyzing publication bias

Publication bias, or the “file drawer problem” (Rosenthal, 1997), refers to the phenomenon that studies reporting favorable results may be more likely to be published than studies with poor or non-findings. Accordingly, the effect sizes and variances from a meta-analysis may be under- or overestimated. The literature gives some advice for testing the presence of publication bias (Field & Gillett, 2010; Ringquist, 2013; Steel et al., 2021). First, a “funnel plot” provides a visual impression as to whether the effect sizes were symmetrically distributed (Duval & Tweedie, 2000). Second, the Egger’s test (Egger et al., 1997) is a statistical test for funnel plot asymmetry. Third, in the case of such indication of asymmetry, the trim-and-fill method calculates the number of additional effect sizes that would be necessary to achieve funnel plot symmetry (Duval & Tweedie, 2000).

RESULTS

Figure 2 shows the results of the coding procedure as a two-mode network of leadership styles and outcome variables. Overall, the network can be conceived of as a nomological network of leadership research in PA scholarship. A measure of association between all moderators (i.e., the higher-order categories in the coding scheme) is presented in online Appendix F. The table shows small to medium associations, which implies that these

characteristics of leadership research are not evenly distributed and combined but co-occur in some patterns. For instance, some leadership styles are more likely to be studied in certain administrative traditions than in others. A closer inspection of the contingency and correlation tables in online Appendices G and H, showing associations between the lower-order coding categories, reveals that, for instance, the study of ethical leadership clusters to some extent in the Anglo-American and East Asian tradition while being slightly underrepresented in the Germanic and Scandinavian tradition. However, such contingencies should be interpreted with caution, as the number of correlation coefficients is still small and an uneven distribution across coding categories is therefore likely.

Meta-analysis of main effects

Table 1 presents the results of the multi-level meta-analysis for the relationship between leadership and beneficial outcomes across all leadership styles. The mean effect size is positive and significant ($\rho = .380$, 95% CI: [0.347; 0.413], $p < .001$), corresponding to a large correlation (Steel et al., 2021). Table 1 also presents the results for detrimental outcomes. The mean effect size is negative and significant ($\rho = -.196$, 95% CI: [-0.254; -0.136], $p < .001$), indicating a medium correlation (Steel et al., 2021).

The Q -test shows substantial heterogeneity in correlations for both kinds of leadership outcomes. For beneficial outcomes ($Q = 336035.0937$, $df = 441$, $p < .001$), there is a medium-high variance between ($\sigma_1 = .182$, $I_b^2 = 46.41\%$) and within samples ($\sigma_2 = .195$, $I_w^2 = 53.32\%$; e.g., Higgins & Thompson, 2002). In the case of detrimental outcomes ($Q = 1732.2603$, $df = 43$,

$p < .001$), the Q -test reveals low variation between ($\sigma_1 = .063$, $I_b^2 = 10.34\%$) and large variation within samples ($\sigma_2 = .184$, $I_w^2 = 89.22\%$). These results provide strong indications of the presence of moderating effects in the leadership-outcome relationship.

Moderator analyses

The moderator analyses follow the framework presented in Figure 1. To ensure sufficient statistical power, only those moderator categories were included for which at least five correlation coefficients were available from primary studies. For reasons of brevity, the presented results are limited to moderations of the relationship between leadership and beneficial outcomes because such outcomes have received far more attention from leadership scholars than detrimental outcomes (Figure 2, online Appendix D).

Leadership style

Table 2 presents the moderator analysis for the dummy-coded leadership style. The overall results confirm that leadership style indeed accounts for heterogeneity in correlation coefficients within and across samples ($Q_M = 43.0561$, $df = 13$, $p < .001$, $\sigma_1^2 = 0.034$, $\sigma_2^2 = 0.034$). As transformational leadership is the most frequently researched style (Figure 2), it serves as a reference category in the meta-regression. Most leadership styles do not differ significantly in their effect sizes from transformational leadership. Exceptions are controlling leadership ($b = -.297$, $p = .0015$) and laissez-faire leadership ($b = -.392$, $p < .001$), which have smaller mean correlations at statistically significant levels. Given the broad focus of this analysis on

TABLE 1 Results of multi-level meta-analysis for overall relationship between leadership and outcomes

	Estimate	95% CI		p	I^2	k	n
		LB	UB				
Beneficial outcomes							
Fixed effect							
Overall effect size (ρ)	0.380***	0.347	0.413	<.001	99.73%	442	2,380,103
Random effects							
Between-sample variance (τ_b^2)	0.033				46.41%		
Within-sample variance (τ_w^2)	0.038				53.32%		
Detrimental outcomes							
Fixed effect							
Overall effect size (ρ)	-0.196***	-0.254	-0.136	<.001	99.56%	44	439,488
Random effects							
Between-sample variance (τ_b^2)	0.004				10.34%		
Within-sample variance (τ_w^2)	0.034				89.22%		

Abbreviations: 95% CI, 95% confidence interval of ρ ; estimate, estimated correlation coefficient (ρ); k , number of correlation coefficients used; LB, lower bound; n , sample size; UB, upper bound.

*** $p < .001$.

TABLE 2 Moderator analysis for leadership style and beneficial outcomes

	Estimate	95% CI		<i>p</i>	<i>k</i>	<i>n</i>
		LB	UB			
Transformational (ref.) ^a	0.373***	0.322	0.421	<.001	159	461,822
Ambidextrous/paradoxical	0.025	-0.205	0.253	.834	8	5562
Change/entrepreneurial	0.068	-0.1	0.232	.427	22	484,713
Controlling/directive	-0.297**	-0.459	-0.117	.002	9	23,829
Empowering/supporting	0.027	-0.081	0.135	.619	53	273,903
Ethical	0.113	-0.026	0.247	.110	30	176,787
Laissez-faire	-0.392***	-0.549	-0.209	<.001	5	8443
LMX	0.027	-0.112	0.165	.704	27	11,428
Participative	0.073	-0.188	0.325	.585	5	4229
Public	-0.165	-0.378	0.065	.158	9	8470
Servant	-0.024	-0.166	0.119	.74	23	8775
Shared	0.003	-0.174	0.18	.975	10	6153
Task-oriented	-0.113	-0.343	0.13	.362	5	4008
Transactional	-0.067 ⁺	-0.145	0.012	.097	40	328,637

Abbreviations: 95% CI, 95% confidence interval of *p*; estimate, estimated correlation coefficient (*p*); *k*, number of correlation coefficients used; LB, lower bound; *n*, sample size; UB, upper bound.

^aCorresponds to intercept in regression model.

****p* < .001. ***p* < .01. +*p* < .1.

beneficial outcomes in aggregate, online Appendix M presents meta-analytic results for all combinations of leadership styles with particular outcomes (e.g., job satisfaction, organizational commitment, public service motivation (PSM) for which five or more correlation coefficients were available from published studies.

Kind of outcome

Table 3 shows the results of the moderator analysis for the kind of leadership outcome (i.e., employee-, job-, management-, group/organization-related). The strength of the leadership-outcome relationship indeed varies across these kinds ($Q_M = 14.1662$, $df = 3$, $p = .0027$, $\sigma_1^2 = 0.036$, $\sigma_2^2 = 0.037$). Employee-related outcomes serve as a reference category because such outcomes have most frequently been studied. While the leadership-outcome link is neither stronger nor weaker for the job- and management-related outcomes, the moderator analysis reveals that correlations are significantly stronger in the case of a group/organization-related outcome ($b = .094$, $p < .001$).

Administrative subfield

Table 4 shows that the strength of the leadership-outcome relationship differs only slightly across various administrative subfields ($Q_M = 7.1855$, $df = 4$, $p = .1264$, $\sigma_1^2 = 0.033$, $\sigma_2^2 = 0.041$). The subfields of law enforcement/military ($b = -.157$, $p = .072$) and education ($b = -.137$, $p = .049$) have marginally weaker correlations

with outcome variables compared to core administrative services, which serve as the reference category in the model.

Administrative tradition

Table 5 shows the results of the moderator analysis for administrative tradition, which is indeed a significant source of heterogeneity in effect sizes ($Q_M = 24.8595$, $df = 7$, $p < .001$, $\sigma_1^2 = 0.028$, $\sigma_2^2 = 0.037$). As most studies have been conducted in Anglo-American countries, this tradition is the reference category. The leadership-outcome correlation is significantly smaller in East Asian countries ($b = -.136$, $p = .022$) and countries in the Soviet tradition ($b = -.129$, $p = .042$) as well as two continental European traditions: the Germanic ($b = -.183$, $p = .001$) and the Scandinavian tradition ($b = -.276$, $p < .001$).

Leadership style and administrative tradition

Given that the moderator analysis for administrative tradition reveals cultural differences in the leadership-outcome relationship, the question arises whether such differences hold for all leadership styles to the same extent. Figure 3 shows the average correlation coefficients for the relationship between different leadership styles and beneficial outcomes for three subgroups (i.e., Anglo-American, continental European, and East Asian). The analysis is limited to these subgroups because

TABLE 3 Moderator analysis for kind of beneficial outcome

	Estimate	95% CI		<i>p</i>	<i>k</i>	<i>n</i>
		LB	UB			
Employee-related (ref.) ^a	0.356***	0.318	0.393	<.001	250	1,386,269
Group-/organization-related	0.094***	0.044	0.144	<.001	137	654,117
Job-related	-0.014	-0.112	0.085	.788	26	27,989
Management-related	0.045	-0.045	0.133	.328	29	311,728

Abbreviations: 95% CI, 95% confidence interval of *p*; estimate, estimated correlation coefficient (*p*); *k*, number of correlation coefficients used; LB, lower bound; *n*, sample size; UB, upper bound.

^aCorresponds to intercept in regression model.

****p* < .001.

TABLE 4 Moderator analysis for administrative subfield and beneficial outcomes

	Estimate	95% CI		<i>p</i>	<i>k</i>	<i>n</i>
		LB	UB			
Core administration (ref.) ^a	0.402***	0.362	0.441	<.001	289	2,258,616
Education	-0.137*	-0.268	-0.001	.049	37	21,885
Health	-0.057	-0.202	0.09	.447	24	15,044
Law enforcement/military	-0.157 ⁺	-0.318	0.014	.072	34	29,649
Social services	0.043	-0.152	0.236	.665	22	3281

Abbreviations: 95% CI, 95% confidence interval of *p*; estimate, estimated correlation coefficient (*p*); *k*, number of correlation coefficients used; LB, lower bound; *n*, sample size; UB, upper bound.

^aCorresponds to intercept in regression model.

****p* < .001. **p* < .05. ⁺*p* < .1.

TABLE 5 Moderator analysis for administrative tradition and beneficial outcomes

	Estimate	95% CI		<i>p</i>	<i>k</i>	<i>n</i>
		LB	UB			
Anglo-American (ref.) ^a	0.460***	0.407	0.51	<.001	143	2,041,610
East Asian	-0.136*	-0.248	-0.02	.022	70	60,521
Germanic	-0.183**	-0.29	-0.071	.001	77	180,139
Islamic	-0.102	-0.36	0.17	.464	5	1126
Napoleonic	0.008	-0.143	0.158	.921	27	38,995
Postcolonial	-0.059	-0.2	0.084	.42	29	6837
Scandinavian	-0.276***	-0.396	-0.146	<.001	42	24,153
Soviet	-0.129*	-0.249	-0.005	.042	39	12,227

Abbreviations: 95% CI, 95% confidence interval of *p*; estimate, estimated correlation coefficient (*p*); *k*, number of correlation coefficients used; LB, lower bound; *n*, sample size; UB, upper bound.

^aCorresponds to intercept in regression model.

****p* < .001. ***p* < .01. **p* < .05.

the number of correlations from other traditions is still too small to warrant inclusion. For the same reason, the various continental European traditions (i.e., Germanic, Napoleonic, Scandinavian) are pooled into a joint group. A noteworthy finding is that the confidence intervals for transformational leadership in the Anglo-American and the continental European subgroup do not overlap, indicating that there is a significant difference between these effect sizes (e.g., Cumming, 2009). The detailed results of the subgroup analyses are presented in online Appendix

I. Online Appendices J and K mirror these analyses for administrative subfields.

Leadership rating

The moderator analysis (Table 6) indicates no significant differences in correlation coefficients depending on the rating method of leadership (i.e., self- vs. other-rated; $Q_M = 2.0622$, $df = 1$, $p = .151$, $\sigma_1^2 = 0.033$, $\sigma_2^2 = 0.039$). It

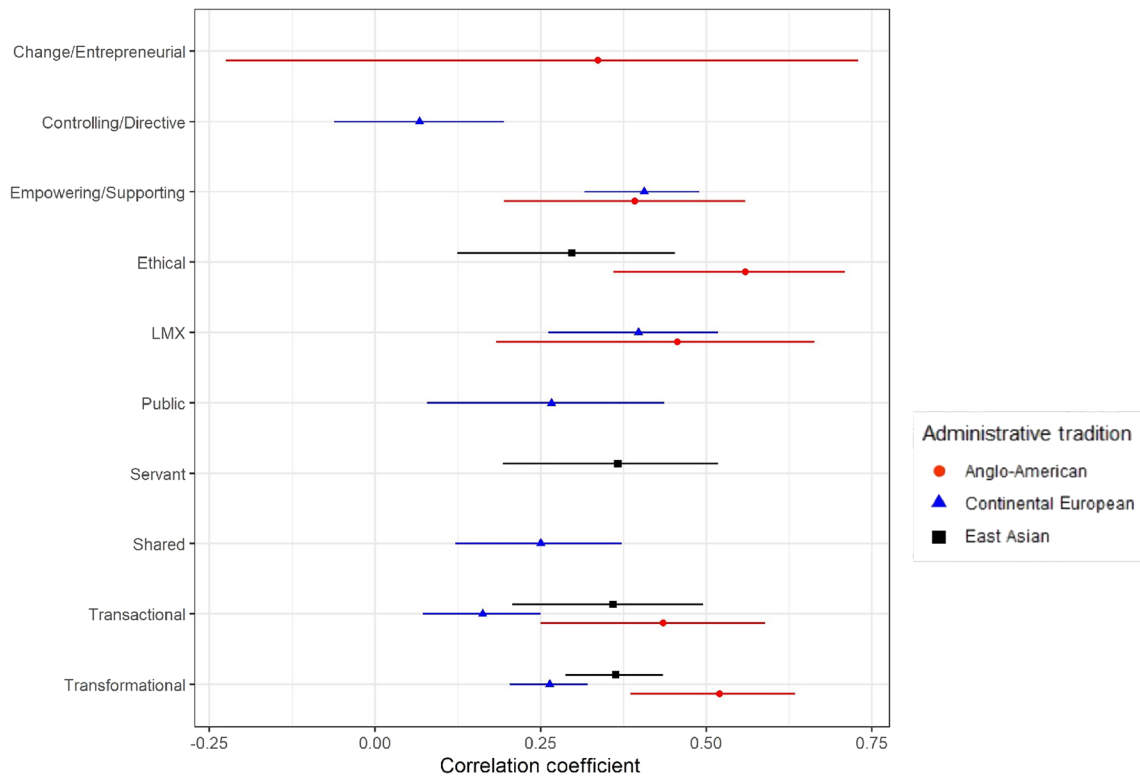


FIGURE 3 Subgroup analyses of leadership-outcome relationship in Anglo-American, continental European, and East Asian tradition

TABLE 6 Moderator analysis for leadership rating

	Estimate	95% CI		p	k	n
		LB	UB			
Others-rated (ref.) ^a	0.385***	0.351	0.417	<.001	412	2,372,604
Self-rated	-0.089	-0.209	0.033	.151	28	6907

Abbreviations: 95% CI, 95% confidence interval of ρ ; estimate, estimated correlation coefficient (ρ); k, number of correlation coefficients used; LB, lower bound; n, sample size; UB, upper bound.

^aCorresponds to intercept in regression model.

*** $p < .001$.

should be noted here that only a vast minority of effects (6.2%) is based on leadership self-ratings (online Appendix G) and that results should therefore be considered with caution.

Data source

Table 7 shows the results of the moderator analysis for the data source (i.e., single source vs. multi-source). The data source indeed explains variance across effect sizes to significant extents ($Q_M = 13.9863$, $df = 1$, $p < .001$, $\sigma_1^2 = 0.031$, $\sigma_2^2 = 0.038$). The moderator analysis reveals that correlations are significantly weaker if independent and dependent variables are based on different data sources ($b = -.152$, $p < .001$).

Research design

Table 8 presents the results of the moderator analysis for the research design (i.e., cross-sectional vs. time-lagged). The strength of the leadership-outcome relationship indeed varies across these designs ($Q_M = 4.8624$, $df = 1$, $p = .0274$, $\sigma_1^2 = 0.033$, $\sigma_2^2 = 0.038$). Correlations are significantly weaker if there is a temporal separation between the measurement of the independent and dependent variable ($b = -.114$, $p = .027$).

Publication bias

According to the visual test method, funnel plot asymmetry for the relationship between leadership and all

TABLE 7 Moderator analysis for data source

	Estimate	95% CI		<i>p</i>	<i>k</i>	<i>n</i>
		LB	UB			
Single source (ref.) ^a	0.393***	0.361	0.425	<.001	390	2,365,681
Multi source	-0.152***	-0.23	-0.073	<.001	52	14,422

Abbreviations: 95% CI, 95% confidence interval of ρ ; estimate, estimated correlation coefficient (ρ); *k*, number of correlation coefficients used; LB, lower bound; *n*, sample size; UB, upper bound.

^aCorresponds to intercept in regression model.

*** $p < .001$.

TABLE 8 Moderator analysis for research design

	Estimate	95% CI		<i>p</i>	<i>k</i>	<i>n</i>
		LB	UB			
Cross-sectional (ref.) ^a	0.389***	0.355	0.421	<.001	404	2,369,111
Time-lagged	-0.114*	-0.213	-0.013	.027	38	10,992

Abbreviations: 95% CI, 95% confidence interval of ρ ; estimate, estimated correlation coefficient (ρ); *k*, number of correlation coefficients used; LB, lower bound; *n*, sample size; UB, upper bound.

^aCorresponds to intercept in regression model.

*** $p < .001$. * $p < .05$.

outcomes (i.e., beneficial and detrimental) is not given (online Appendix L). An insignificant result of the Egger's test (Egger et al., 1997; $z = -1.082$, $p = .279$) supports the assumption of funnel plot symmetry. Accordingly, these methods show no indication of publication bias, and no further analysis with the trim-and-fill method (Duval & Tweedie, 2000) is necessary.

DISCUSSION

This meta-analysis has extracted the nomological network of quantitative leadership studies and examined the kind and strength of relationships between leadership and outcome variables. Even more importantly, it explores conceptual, contextual, and methodological boundary conditions under which leadership is associated with outcomes. The results have several implications for theory, research, and practice. *First*, the meta-analysis shows that leadership does make a difference in the public sector. The meta-analytic results across 151 studies establish a strong association of leadership with beneficial outcomes and a moderate relationship with detrimental outcomes. The job demands-resources (JD-R) model may offer a possible explanation for the different strengths of these relationships (Tummers & Bakker, 2021): Arguably, positive forms of leadership are a job resource for employees and thus initiate a motivational cycle that yields various beneficial outcomes (e.g., individual performance, organizational commitment). In contrast, detrimental outcomes are the result of a loss cycle. Job resources interfere with this loss cycle only indirectly through moderating effects, which may explain why the association of leadership with detrimental outcomes is weaker than the more direct

motivational effects on beneficial outcomes. However, this interpretation needs further elaboration, as it suggests a general tendency that may not apply to specific leadership outcomes (some of which resonate better with the JD-R model than others).

It is also important to note that, in terms of valence, scholarship has covered far less variation in the independent than in the dependent variable in the leadership equation. While outcomes reflect a broad range of valence (yet with a bias toward beneficial at the expense of detrimental outcomes), leadership itself has predominantly been studied in positive forms. The whole field of "bad" leadership, such as "abusive supervision" (Luu, 2018a) or "toxic leadership" (Williams, 2018), has only very rarely been studied in PA and could thus not be included in this meta-analysis. It is unclear whether negative forms of leadership simply yield the reverse effects of positive leadership in similar strengths. As negative leadership may do more harm than positive leadership does good, more research on the "dark side" of leadership is needed.

Second, even on the "bright" side of leadership, there is variance in the difference that leadership makes, depending on the style of leadership. The results illustrate that transformational leadership is the most frequently studied style (Chapman et al., 2016; Vogel & Masal, 2015) and support the community in setting this priority, as transformational leaders indeed make a strong and consistent positive difference in public organizations. Unsurprisingly, transformational leadership shows a significantly stronger relationship with positive outcomes than laissez-faire leadership, which has repeatedly been demonstrated to be largely ineffective (Yukl & Gardner, 2020). The same applies to controlling

leadership (Hattke et al., 2018; Vermeeren et al., 2014). This finding deserves attention, as public organizations are often characterized by a high density of rules and regulations, which may require or invite controlling behaviors on the part of leaders (Tummers & Knies, 2016).

The moderator analysis also shows that there are several other leadership styles that have no weaker associations with beneficial outcomes than transformational leadership. This is good news for practitioners in human resource (HR) management, as not all leadership styles can be easily developed across different people and settings. Neither do actual and prospect leaders bring all qualifications for any leadership style to their careers, nor do all settings provide the same facilitating conditions for various leadership styles to emerge. A broad range of leadership styles thus helps HR practitioners to adjust leadership trainings to the available talents and given settings. Public leaders may feel encouraged that there are many ways to become an effective leader. Within the range of the analyzed styles positively associated with beneficial outcomes, it is not particularly important *how* they lead — as long as they *do* lead. Public leadership scholars have therefore good reasons to further expand on the nomological network of leadership and intensify research on hitherto understudied leadership styles, such as shared, authentic, or network leadership. This would also facilitate conversations in the field with general leadership studies (Ospina, 2017), where the conceptual variety is still substantially larger (Van Wart, 2003, 2013a).

It should be noted, however, that several leadership styles are not so different after all but show considerable conceptual and empirical overlaps (e.g., Banks et al., 2016; Hoch et al., 2016). Analyzing the combined effects of different aspects of leadership, rather than to focus on specific styles, might be an interesting research stream. While the great variety of leadership styles reflects important and distinct patterns of leadership behaviors, a stronger combinative and configurational thinking about these styles, culminating in a new “full-range” theory of leadership, is promising for future research and practice (Anderson & Sun, 2017). Along these lines, Günzel-Jensen et al. (2018) investigate interactions between transformational, transactional, and empowering leadership in the prediction of public employees’ innovative behavior, advancing the understanding of how different styles jointly affect outcomes. This is a particularly promising avenue for public sector research because stakeholders’ expectations toward public leaders are often conflicting and thus difficult to address within the scope of a specific style.

Third, the results show that the leadership-outcome relationship in the public sector is stronger for group and organization-related outcomes than for outcomes related to the individual employee. This is a surprising finding, given that leadership has been defined as an interactional process of influencing others, which makes followers the primary target and defining instance of leadership (Yukl &

Gardner, 2020). A possible explanation for this finding is that there may be more noise in measures of individual outcomes than in aggregated higher-level outcomes (e.g., Ostroff, 1993; Song & Meier, 2018). Considering the adequate analytic level of analysis (i.e., individuals, groups, etc.) is a lasting debate in several subfields of PA, especially in performance measurement (e.g., Andersen et al., 2016). Another possible explanation is that administrative leadership resonates not only with internal stakeholders (i.e., followers) but also with external audiences; there is hence a second and additional route through which it affects organization-level outcomes. For example, leadership may build political or community support, which is likely to translate into organizational performance. This issue is left for future research, as previous scholarship has only rarely examined outcomes of administrative leadership beyond the organizational sphere. Given the dedication of public organizations to the political and societal will, more research on the impact of administrative leadership on outcomes at these levels is overdue.

Fourth, context matters for public leadership, but different contextual factors matter to different extents. Surprisingly, the correlations of leadership styles with outcomes vary only slightly across administrative subfields. Given that leadership studies often stress their origin in military leadership (Yukl & Gardner, 2020), it is an interesting finding that the leadership-outcome link is slightly weaker in this field compared to core administration. The scope of this meta-analysis may have shifted the focus to leadership styles that are more interesting to the PA audience but less effective in “command-and-control” organizations (Hattke et al., 2018; McCann & Pigeau, 2000), such as the police and the military, where more authoritarian leadership styles may be prevalent. Compared to core administration, the leadership-outcome link is also weaker in educational organizations. The work-related norms and values of teachers are likely to foster job autonomy and peer-control which should make members less receptive toward vertical leadership and thus may gradually substitute for leadership. However, these differences in effect sizes occur only at marginal levels of statistical significance, such that the overall conclusion can be drawn that leadership does not substantially differ between administrative subfields. Accordingly, leadership scholars and practitioners who are specialized in a particular subfield should feel encouraged to engage in exchange not only within but also across their subcommunities.

Fifth, the context of administrative tradition does matter to leadership. This result speaks to scholars who call for more consideration of the context in PA, both in leadership studies (Ospina, 2017; Vogel & Masal, 2015) and beyond (Meier et al., 2017; O’Toole & Meier, 2014). It is consistent with a recent study by Raudla et al. (2021), who also find that the administrative tradition explains more variance in attitudes and behaviors than the policy field. While leadership is nowhere unrelated to outcomes, the strength of this relationship varies across different traditions. In particular, leadership in public sectors in the

Anglo-American tradition shows significantly stronger associations with beneficial outcomes than leadership in the Germanic, Scandinavian, East Asian, and Soviet traditions. Both the Scandinavian and, even more so, the Germanic countries are representatives of the continental European rule-of-law tradition with a high density of rules and regulations (Kuhlmann & Wollmann, 2019; Painter & Peters, 2010). The high degree of formalization in these administrative systems may neutralize the effects of leadership (Kerr & Jermier, 1978; Muchiri & Cooksey, 2011) as it is likely to leave less discretion for both leaders and followers than elsewhere. Furthermore, a stronger emphasis on legality in the continental European tradition may push leaders more toward rule-following behavior (Tummers & Knies, 2016) and controlling leadership (Hattke et al., 2018), which have been shown to be less effective than other styles of leadership. In contrast, the administrative culture in Anglo-American countries are more pragmatic and flexible, and education and training place more emphasis on managerial and leadership skills. This context may also set a better stage for charismatic and visionary leaders and may thus explain why transformational leadership shows a significantly stronger association with outcomes in Anglo-American than in continental European countries. For the same reason, transformational and empowering leadership are also more likely to emerge in Anglo-American countries in the first place. This is suggested by frequent combinations of these styles with the Anglo-American tradition in the contingency table of coding categories (online Appendix G).

It should be noted here that administrative tradition is a rather broad contextual factor. Even within the same tradition, there may be considerable differences in administrative systems and cultures between countries. However, a post-hoc analysis of cross-country differences in the leadership-outcome link within the main traditions (not reported for reasons of brevity) reveals that the country moderates this relationship only in a few cases. Differences in leadership tend to be stronger between administrative traditions than within these traditions, which supports previous studies demonstrating the relevance of administrative tradition as a contextual factor (Mohr et al., 2021; Raudla et al., 2021; Van der Wal et al., 2021).

Apart from institutional characteristics of countries from different administrative traditions, broader cultural values and beliefs might provide additional explanations for variations in the leadership-outcome relationship. According to the GLOBE study (House et al., 2004), many European countries are characterized by lower power distance, more collectivism, and less leader-centrism compared to Anglo-American countries. In particular, this applies to Scandinavia, where leaders show more team-oriented and participative behaviors that strongly focus on involvement and collaboration to achieve common goals. The frequent combination of participative leadership with the Scandinavian tradition points in this direction (online Appendix G). In contrast, Anglo-American

countries are characterized by higher levels of individualism and performance orientation (Hofstede, 2011; House et al., 2004), which might set the better stage for more leader-centric leadership styles such as transformational leadership. In a similar vein, differences in leadership effectiveness between East Asian and Anglo-American countries may have similar reasons, as Asian countries score high on collectivism (Hofstede, 2011; House et al., 2004). As leadership is still predominantly studied as an individualistic, leader-centric concept, it is likely to resonate better with individualistic than with collectivistic cultures. To conclude, leadership interacts with broader traditions and cultures, and further explorations of this interaction set a rich agenda for further research.

Sixth, the results make a case for more robust research designs in public leadership scholarship. A moderator analysis reveals significant differences in the strength of the leadership-outcome relationship between the use of single versus multiple data sources, as the overall correlation is significantly smaller when authors used different sources for the measurement of the independent and dependent variables. Furthermore, the analysis indicates a moderating effect of the research design, with the leadership-outcome relationship being significantly stronger in cross-sectional than in time-lagged designs. Both findings speak to concerns of common method bias (George & Pandey, 2017; Meier & O'Toole, 2013), as the use of multiple data sources and the temporal separation of independent and dependent variables have been suggested as remedies to this issue (Podsakoff et al., 2003). The results show that once scholars follow this advice, the correlations between leadership and outcomes are significantly and substantially smaller. Accordingly, they help to assess the extent to which common method bias inflates the results of leadership studies by sourcing independent and dependent variables from the same respondents and measuring them at the same time. As this is still often the case, scholarship is likely to overestimate the effects of leadership in the public sector. More robust research designs are thus urgently needed in order to deliver unbiased results which, in turn, foster theory development. Leadership scholars in the public sector are thus encouraged to combine different data sources and to apply research designs other than cross-sectional designs. It is remarkable and alarming that coding for truly longitudinal designs and experimental research was not possible in this meta-analysis due to an insufficient number of primary studies complying with these methodological requirements.

Limitations

As in any research, this study is not without limitations. First, it is worth repeating that meta-analysis in PA research is commonly a correlative technique (George et al., 2021; Homberg et al., 2015) that does not allow for causal inference. Referring to correlates as leadership outcomes have been a conceptual rather than

methodological choice, as previous scholarship has predominantly theorized these variables as outcomes of leadership. However, given the primacy of survey research in leadership studies, a high share of the meta-analyzed studies are limited in establishing causal relationships. Second, the coding in this meta-analysis has been adjusted to the current developmental stage of leadership research in the public sector. While the field has grown to a stage that warrants the first meta-analysis, the number of reported effect sizes is still too limited to apply more fine-grained coding categories. An example is different conceptualizations and measurements of the same leadership styles coded into the same category. This applies to transformational leadership, which is conceptualized in different breadth and measured with various scales. If research on transformational leadership in the public sector continues to grow, future meta-analyses will be able to fill a more fine-grained coding scheme with a sufficient number of effect sizes from primary studies. Third, many more contextual factors, beyond those considered in this study, are likely to account for variance in the leadership-outcome link (e.g., organizational size and hierarchy). Before calling for the inclusion of such moderators in meta-analyses, however, it needs to be considered that such information is often not available from primary studies. Fourth, although the funnel plot technique provides statistical indications for the absence of publication bias, a more reliable approach would be to include studies that have not been published yet.

CONCLUSION

Leadership continues to attract considerable scholarly and practical attention in PA (Chapman et al., 2016; Hart & Tummors, 2019; Van Wart, 2013a; Vogel & Masal, 2015). This meta-analysis shows that leadership deserves this (and even more) attention, as it is positively and largely consistently related to intended results across different styles, outcomes, and contexts. The results help to manage expectations about what outcomes to expect from which kind of leadership and under which contextual conditions. As the coverage of single studies is too limited to provide this information, aggregating findings from many studies at a meta-level is a step forward in this direction, although this approach comes at the expense of richness of details. Given the overall result that different leadership styles seem to work well in different public sector settings, scholarship and practice are invited to further broaden the scope of leadership research and to expand the nomological network toward others and combined approaches (Crosby & Bryson, 2018). The findings of this meta-analysis also show how context matters for leadership in the public sector and warns against overgeneralizations, as the effectiveness of leadership depends on the administrative tradition and, yet to a much smaller extent, on the administrative subfield. Conducting leadership research more evenly across these domains, and

accounting for further contextual influences, is encouraged for future studies and meta-analyses thereof. Additionally, this meta-analysis underscores the importance of methodological choices when conducting leadership research. Findings echo concerns of common method bias and thus encourage more diversity in research designs and data sources.

ACKNOWLEDGMENT

Open Access funding enabled and organized by Projekt DEAL.

ORCID

Leonie Backhaus  <https://orcid.org/0000-0002-5372-5617>
Rick Vogel  <https://orcid.org/0000-0002-1361-5085>

REFERENCES

- Ancarani, Alessandro, Francesco Arcidiacono, Carmela Di Mauro, and Maria Daniela Giammanco. 2020. "Promoting Work Engagement in Public Administrations: The Role of Middle Managers' Leadership." *Public Management Review* 23(8): 1234–1263.
- Andersen, Lotte Bøgh, Andreas Boesen, and Lene Holm Pedersen. 2016. "Performance in Public Organizations: Clarifying the Conceptual Space." *Public Administration Review* 76(6): 852–62.
- Anderson, Marc H., and Peter Y. T. Sun. 2017. "Reviewing Leadership Styles: Overlaps and the Need for a New 'Full-Range' Theory." *International Journal of Management Reviews* 19(1): 76–96.
- Asencio, Hugo. 2016. "Leadership, Trust, and Job Satisfaction in the Public Sector: A Study of US Federal Employees." *International Review of Public Administration* 21(3): 250–67.
- Backhaus, Leonie, Artur Reuber, Dominik Vogel, and Rick Vogel. 2021. "Giving Sense about Paradoxes: Paradoxical Leadership in the Public Sector." *Public Management Review* forthcoming.
- Banks, George C., Kelly Davis McCauley, William L. Gardner, and Courtney E. Guler. 2016. "A Meta-Analytic Review of Authentic and Transformational Leadership: A Test for Redundancy." *The Leadership Quarterly* 27(4): 634–52.
- Bauwens, Robin, Mieke Audenaert, and Adeliën Decramer. 2019. "Fostering Societal Impact and Job Satisfaction: The Role of Performance Management and Leader-Member Exchange." *Public Management Review* 21(10): 1486–515.
- 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.
- Caillier, James Gerard. 2014. "Toward a Better Understanding of the Relationship between Transformational Leadership, Public Service Motivation, Mission Valence, and Employee Performance: A Preliminary Study." *Public Personnel Management* 43(2): 218–39.
- Cantarelli, Paola, Paolo Belardinelli, and Nicola Belle. 2016. "A Meta-Analysis of Job Satisfaction Correlates in the Public Administration Literature." *Review of Public Personnel Administration* 36(2): 115–44.
- Chapman, Carrie, Heather Getha-Taylor, Maja Husar Holmes, Willow S. Jacobson, Ricardo S. Morse, and Jessica E. Sowa. 2016. "How Public Service Leadership Is Studied: An Examination of a Quarter Century of Scholarship." *Public Administration* 94(1): 111–28.
- Crosby, Barbara C., and John M. Bryson. 2018. "Why Leadership of Public Leadership Research Matters: And What to Do about it." *Public Management Review* 20(9): 1265–86.
- Cumming, Geoff. 2009. "Inference by Eye: Reading the Overlap of Independent Confidence Intervals." *Statistics in Medicine* 28(2): 205–20.
- Duval, Sue, and Richard Tweedie. 2000. "Trim and Fill: A Simple Funnel-Plot-Based Method of Testing and Adjusting for Publication Bias in Meta-Analysis." *Biometrics* 56: 455–63.

- Egger, Matthias, George Davey Smith, Martin Schneider, and Christoph Minder. 1997. "Bias in Meta-Analysis Detected by a Simple, Graphical Test." *British Medical Journal* 315(7109): 629–34.
- Elkomy, Shima, Zahra Murad, and Veronica Veleanu. 2020. "Does Leadership Matter for Healthcare Service Quality? Evidence from NHS England." *International Public Management Journal*.
- Fernandez, Sergio, Yoon J. Cho, and James Perry. 2010. "Exploring the Link between Integrated Leadership and Public Sector Performance." *Leadership Quarterly* 21(2): 308–23.
- Field, Andy P., and Raphael Gillett. 2010. "How to Do a Meta-Analysis." *The British Journal of Mathematical and Statistical Psychology* 63(Pt 3): 665–94.
- George, Bert, and Sanjay K. Pandey. 2017. "We Know the Yin—But Where Is the Yang? Toward a Balanced Approach on Common Source Bias in Public Administration Scholarship." *Review of Public Personnel Administration* 37(2): 245–70.
- George, Bert, Sanjay K. Pandey, Bram Steijn, Adeliën Decramer, and Mieke Audenaert. 2021. "Red Tape, Organizational Performance, and Employee Outcomes: Meta-Analysis, Meta-Regression, and Research Agenda." *Public Administration Review* 81(4): 638–65.
- Geyskens, Inge, Rekha Krishnan, Jan-Benedict E. M. Steenkamp, and Paulo V. Cunha. 2008. "A Review and Evaluation of Meta-Analysis Practices in Management Research." *Journal of Management* 35(2): 393–419.
- Günzel-Jensen, Franziska, Jesper Rosenberg Hansen, Mads Leth Felsager Jacobsen, and Jesper Wulff. 2018. "A Two-Pronged Approach? Combined Leadership Styles and Innovative Behavior." *International Journal of Public Administration* 41(12): 957–70.
- Hart, Paul, and Lars Tummors. 2019. *Understanding Public Leadership*, 2nd ed. London: Red Globe Press.
- Hassan, Shahidul, Leisha DeHart-Davis, and Zhongnan Jiang. 2019. "How Empowering Leadership Reduces Employee Silence in Public Organizations." *Public Administration* 97(1): 116–31.
- Hassan, Shahidul, Jongsoo Park, and Jos C. N. Raadschelders. 2019. "Taking a Closer Look at the Empowerment-Performance Relationship: Evidence from Law Enforcement Organizations." *Public Administration Review* 79(3): 427–38.
- Hassan, Shahidul, Bradley E. Wright, and Gary Yukl. 2014. "Does Ethical Leadership Matter in Government? Effects on Organizational Commitment, Absenteeism, and Willingness to Report Ethical Problems." *Public Administration Review* 74(3): 333–43.
- Hattke, Fabian, Rick Vogel, and Judith Znanewitz. 2018. "Satisfied with Red Tape? Leadership, Civic Duty, and Career Intention in the Military." *Public Management Review* 20(4): 563–86.
- Higgins, Julian P. T., and Simon G. Thompson. 2002. "Quantifying Heterogeneity in a Meta-Analysis." *Statistics in Medicine* 21(11): 1539–58.
- Hoch, Julia E., William H. Bommer, James H. Dulebohn, and Wu. Dongyuan. 2016. "Do Ethical, Authentic, and Servant Leadership Explain Variance above and beyond Transformational Leadership? A Meta-Analysis." *Journal of Management* 44(2): 501–29.
- Hofstede, Geert. 2011. "Dimensionalizing Cultures: The Hofstede Model in Context." *Online Readings in Psychology and Culture* 2(1): 2307.
- Homberg, Fabian, Dermot McCarthy, and Vurain Tabvuma. 2015. "A Meta-Analysis of the Relationship between Public Service Motivation and Job Satisfaction." *Public Administration Review* 75(5): 711–22.
- House, Robert J., Paul J. Hanges, Mansour Javidan, Peter W. Dorfman, and Vipin Gupta, eds. 2004. *Culture, Leadership, and Organizations. The GLOBE Study of 62 Societies*. Thousand Oaks, CA: Sage.
- Jacobsen, Christian Bøtcher, and Lotte Bøgh 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.
- Jacobsen, Christian Bøtcher, Lotte Bøgh Andersen, Anne Bøllingtoft, and Tine Louise Mundbjerg Eriksen. 2021. "Can Leadership Training Improve Organizational Effectiveness? Evidence from a Randomized Field Experiment on Transformational and Transactional Leadership." *Public Administration Review* 82(1): 117–31.
- Jacobsen, Christian Bøtcher, and Camilla Denager Staniok. 2020. "Sharing the Fire? The Moderating Role of Transformational Leadership Congruence on the Relationship between Managers' and Employees' Organizational Commitment." *International Public Management Journal* 23(4): 564–88.
- Jakobsen, Morten, and Rasmus Jensen. 2015. "Common Method Bias in Public Management Studies." *International Public Management Journal* 18(1): 3–30.
- Kerr, Steven, and John M. Jermier. 1978. "Substitutes for Leadership—Their Meaning and Measurement." *Organizational Behavior and Human Performance* 22(3): 375–403.
- Klijn, Erik Hans, Noortje de Boer, and Jasper Eshuis. 2020. "Leading Frontline Enforcers: How Supervisors' Leadership Style Impacts Inspectors' Enforcement Style." *Public Management Review* 24(3): 398–417.
- Konstantopoulos, Spyros. 2011. "Fixed Effects and Variance Components Estimation in Three-Level Meta-Analysis." *Research Synthesis Methods* 2(1): 61–76.
- Kroll, Alexander, and Dominik Vogel. 2014. "The PSM-Leadership Fit: A Model of Performance Information Use." *Public Administration* 92(4): 974–91.
- Kuhlmann, Sabine, and Hellmut Wollmann. 2019. *Introduction to Comparative Public Administration: Administrative Systems and Reform in Europe*, 2nd ed. Cheltenham: Edward Elgar.
- Langan, Dean, Julian P. T. Higgins, Dan Jackson, Jack Bowden, Areti Angeliki Veroniki, Evangelos Kontopantelis, Wolfgang Viechtbauer, and Mark Simmonds. 2019. "A Comparison of Heterogeneity Variance Estimators in Simulated Random-Effects Meta-Analyses." *Research Synthesis Methods* 10(1): 83–98.
- Luu, Trong Tuan. 2018a. "Discretionary HR Practices and Proactive Work Behaviour: The Mediation Role of Affective Commitment and the Moderation Roles of PSM and Abusive Supervision." *Public Management Review* 20(6): 789–823.
- Luu, Trong Tuan. 2018b. "Behind the Influence of Job Crafting on Citizen Value co-Creation with the Public Organization: Joint Effects of Paternalistic Leadership and Public Service Motivation." *Public Management Review* 20(10): 1533–61.
- Masal, Doris, and Rick Vogel. 2016. "Leadership, Use of Performance Information, and Job Satisfaction: Evidence from Police Services." *International Public Management Journal* 19(2): 208–34.
- McCann, Carol, and Ross Pigeau. 2000. *The Human in Command Exploring the Modern Military Experience*. New York: Springer.
- Meier, Kenneth J., and Laurence J. O'Toole. 2013. "I Think (I am Doing Well), Therefore I am: Assessing the Validity of Administrators' Self-Assessments of Performance." *International Public Management Journal* 16(1): 1–27.
- Meier, Kenneth J., Amanda Rutherford, and Claudia N. Avellaneda, eds. 2017. *Comparative Public Management: Why National, Environmental, and Organizational Context Matters*. Washington, DC: Georgetown University Press.
- Miao, Qing, Alexander Newman, Gary Schwarz, and Brian Cooper. 2018. "How Leadership and Public Service Motivation Enhance Innovative Behavior." *Public Administration Review* 78(1): 71–81.
- Mohr, Zachary T., Ringa Raudla, and James W. Douglas. 2021. "Comparing Cost Accounting Use across European Countries: The Role of Administrative Traditions, NPM Instruments, and Fiscal Stress." *Public Administration Review* 81(2): 299–307.
- Muchiri, Michael K., and Ray W. Cooksey. 2011. "Examining the Effects of Substitutes for Leadership on Performance Outcomes." *Leadership & Organization Development Journal* 32(8): 817–36.
- Muterera, Jonathan, David Hemsworth, Anahita Baregheh, and Blanca Rosa Garcia-Rivera. 2018. "The Leader-Follower Dyad: The Link between Leader and Follower Perceptions of Transformational Leadership and its Impact on Job Satisfaction and Organizational Performance." *International Public Management Journal* 21(1): 131–62.
- Ospina, Sonia M. 2017. "Collective Leadership and Context in Public Administration: Bridging Public Leadership Research and Leadership Studies." *Public Administration Review* 77(2): 275–87.
- Ostroff, Cheri. 1993. "Comparing Correlations Based on Individual-Level and Aggregated Data." *Journal of Applied Psychology* 78(4): 569–82.

- O'Toole, Laurence J., Jr., and Kenneth J. Meier. 2014. "Public Management, Context, and Performance: In Quest of a more General Theory." *Journal of Public Administration Research and Theory* 25(1): 237–56.
- Painter, Martin, and B. Guy Peters. 2010. "Administrative Traditions in Comparative Perspective: Families, Groups and Hybrids." In *Tradition and Public Administration* 19–30. Hampshire: Palgrave Macmillan.
- Park, Sanghee. 2020. "Gender and Performance in Public Organizations: A Research Synthesis and Research Agenda." *Public Management Review* 23(6): 929–48.
- Park, Sung Min, and Hal G. Rainey. 2008. "Leadership and Public Service Motivation in US Federal Agencies." *International Public Management Journal* 11(1): 109–42.
- Podsakoff, Philip M., Scott B. MacKenzie, Jeong-Yeon Lee, and Nathan P. Podsakoff. 2003. "Common Method Biases in Behavioral Research: A Critical Review of the Literature and Recommended Remedies." *Journal of Applied Psychology* 88(5): 879–903.
- Raudla, Ringa, James W. Douglas, and Jason H. Windett. 2021. "Comparative Analysis of Financial Bureaucracies of Europe: The Influence of Policy Sector Vs Administrative Tradition on the Attitudes of Finance Officials." *Journal of Comparative Policy Analysis* 23(4): 393–415.
- Ringquist, Evan. 2013. *Meta-Analysis for Public Management and Policy*. New York: Jossey-Bass.
- Rosenthal, Robert. 1997. "The File Drawer Problem and Tolerance for Null Results." *Psychological Bulletin* 86(3): 638–41.
- 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.
- Schyns, Birgit, and Jan Schilling. 2013. "How Bad Are the Effects of Bad Leaders? A Meta-Analysis of Destructive Leadership and its Outcomes." *The Leadership Quarterly* 24(1): 138–58.
- Slyke, Van, M. David, and Robert W. Alexander. 2006. "Public Service Leadership: Opportunities for Clarity and Coherence." *The American Review of Public Administration* 36(4): 362–74.
- Song, Miyeon, and Kenneth J. Meier. 2018. "Citizen Satisfaction and the Kaleidoscope of Government Performance: How Multiple Stakeholders See Government Performance." *Journal of Public Administration Research and Theory* 28(4): 489–505.
- Steel, Piers, Sjoerd Beugelsdijk, and Herman Aguinis. 2021. "The Anatomy of an Award-Winning Meta-Analysis: Recommendations for Authors, Reviewers, and Readers of Meta-Analytic Reviews." *Journal of International Business Studies* 52(1): 23–44.
- Sun, Rusi, and Weijie Wang. 2017. "Transformational Leadership, Employee Turnover Intention, and Actual Voluntary Turnover in Public Organizations." *Public Management Review* 19(8): 1124–41.
- Tavares, Gustavo M., Filipe Sobral, and Bradley E. Wright. 2021. "Commitment to Public Values, Charismatic Leadership Attributions, and Employee Turnover in Street-Level Bureaucracies." *Journal of Public Administration Research and Theory* 31(3): 542–60.
- Tummers, Lars G., and Arnold B. Bakker. 2021. "Leadership and Job Demands-Resources Theory: A Systematic Review." *Frontiers in Psychology* 12: 722080.
- Tummers, Lars G., and Eva Knies. 2013. "Leadership and Meaningful Work in the Public Sector." *Public Administration Review* 73(6): 859–68.
- Tummers, Lars, and Eva Knies. 2016. "Measuring Public Leadership: Developing Scales for Four Key Public Leadership Roles." *Public Administration* 94(2): 433–51.
- Tummers, Lars, Bram Steijn, Barbara Nevicka, and Madelon Heerema. 2018. "The Effects of Leadership and Job Autonomy on Vitality: Survey and Experimental Evidence." *Review of Public Personnel Administration* 38(3): 355–77.
- Umans, Timurs, Elin Smith, William Andersson, and William Planken. 2018. "Top Management Teams' Shared Leadership and Ambidexterity: The Role of Management Control Systems." *International Review of Administrative Sciences* 86(3): 444–62.
- Van der Wal, Z., A. Mussagulovala, and C. A. Chen. 2021. "Path-Dependent Public Servants: Comparing the Influence of Traditions on Administrative Behavior in Developing Asia." *Public Administration Review* 81(2): 308–20.
- Van Wart, Montgomery. 2003. "Public-Sector Leadership Theory: An Assessment." *Public Administration Review* 63(2): 214–28.
- Van Wart, Montgomery. 2013a. "Administrative Leadership Theory: A Reassessment After 10 Years." *Public Administration* 91(3): 521–43.
- Van Wart, Montgomery. 2013b. "Lessons from Leadership Theory and the Contemporary Challenges of Leaders." *Public Administration Review* 73(4): 553–65.
- Van Wart, Montgomery, A. Hondeghem, E. Schwella, and Vivien E. Nice. 2015. *Leadership and Culture*. UK: Palgrave Macmillan.
- Vermeeren, Brenda, Ben Kuipers, and Bram Steijn. 2014. "Does Leadership Style Make a Difference? Linking HRM, Job Satisfaction, and Organizational Performance." *Review of Public Personnel Administration* 34(2): 174–95.
- Viechtbauer, Wolfgang. 2010. "Conducting Meta-Analyses in R with the Metafor Package." *Journal of Statistical Software* 36: 1–48.
- Vogel, Dominik, and Alexander Kroll. 2019. "Agreeing to Disagree? Explaining Self-Other Disagreement on Leadership Behaviour." *Public Management Review* 21(12): 1867–92.
- 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.
- Vogel, Dominik, Artur Reuber, and Rick Vogel. 2020. "Developing a Short Scale to Assess Public Leadership." *Public Administration* 98(4): 958–73.
- Williams, Kenneth R. 2018. "Toxic Leadership in Defense and Federal Workplaces: Sabotaging the Mission and Innovation." *International Journal of Public Leadership* 14(3): 179–98.
- 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.
- Wright, Bradley E., and Sanjay K. Pandey. 2010. "Transformational Leadership in the Public Sector: Does Structure Matter?" *Journal of Public Administration Research and Theory* 20(1): 75–89.
- Yukl, Gary A., and William L. Gardner. 2020. *Leadership in Organizations*, 9th ed. Harlow: Pearson.

AUTHOR BIOGRAPHIES

Leonie Backhaus is a research associate at the Chair of Public Management at Universität Hamburg (Germany). Her research interests are leadership in the public sector, leadership and employee development, and motivational, behavioral and performance-related employee outcomes.

Email: leonie.backhaus@uni-hamburg.de

Rick Vogel is a full professor of public management at Universität Hamburg (Germany). His current research interests are public sector leadership, mixed governance, and institutional change in the public sector.

Email: rick.vogel@uni-hamburg.de

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.

How to cite this article: Backhaus, Leonie, and Rick Vogel. 2022. "Leadership in the Public Sector: A Meta-Analysis of Styles, Outcomes, Contexts, and Methods." *Public Administration Review* 82(6): 986–1003. <https://doi.org/10.1111/puar.13516>