

Vogel, Dominik; Reuber, Artur; Vogel, Rick

Article — Published Version

## Developing a short scale to assess public leadership

Public Administration

**Provided in Cooperation with:**

John Wiley & Sons

*Suggested Citation:* Vogel, Dominik; Reuber, Artur; Vogel, Rick (2020) : Developing a short scale to assess public leadership, Public Administration, ISSN 1467-9299, Blackwell Publishing Ltd, Oxford, UK, Vol. 98, Iss. 4, pp. 958-973, <https://doi.org/10.1111/padm.12665>

This Version is available at:

<https://hdl.handle.net/10419/230244>

**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/4.0/>

# Developing a short scale to assess public leadership

Dominik Vogel  | Artur Reuber | Rick Vogel 

Department of Socioeconomics, University of Hamburg, Germany

## Correspondence

Dominik Vogel, Department of Socioeconomics, University of Hamburg, Von-Melle-Park 9, 20146 Hamburg, Germany.  
Email: dominik.vogel-2@uni-hamburg.de

## Abstract

Tummers and Knies (2016) have recently introduced a 21-item scale for the measurement of public leadership to the burgeoning field of leadership research in public administration. However, due to restrictions in survey length and response time, scholars often face practical difficulties when adopting measurement scales of such length. In many subfields of public administration, this results in a proliferation of ad hoc measures of unknown validity, which impedes scholarly progress. The goal of the present study is to develop a short form of the public leadership scale. We build on data from a two-wave study in the German public sector and follow a step-by-step scale reduction procedure. The result is a reliable and valid 11-item scale of public leadership for utilization in public administration research. Since a short scale allows researchers to include additional measures of other constructs, it facilitates the exploration of the nomological network of public leadership.

## 1 | INTRODUCTION

Scholarship in public administration has recently paid growing attention to the phenomenon of leadership (e.g., van Wart 2013; Vogel and Masal 2015; Crosby and Bryson 2018). Leadership is generally defined as ‘the process of influencing others to understand and agree about what needs to be done and how to do it, and the process of facilitating individual and collective efforts to accomplish shared objectives’ (Yukl 2006, p. 8). Given the plurality and ambiguity of organizational goals in the public sector (Rainey and Jung 2015), the need to align these goals towards

---

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2020 The Author. *Public Administration* published by John Wiley & Sons Ltd.

public values and the common good and to motivate organizational members accordingly is evident. While the important role of leaders in this process has been widely acknowledged (e.g., Stazyk and Davis 2020), it is less clear what kind of leadership emerges in the public sector and proves to be effective in this context. At the risk of falling victim to public idiosyncrasies (Perry 2016; Ospina 2017), scholars have argued that the public sector is a unique setting in which elements of 'publicness' provide a distinctive context of leadership (Lambright and Quinn 2011; Vogel and Masal 2015; Tummers and Knies 2016). This gives rise to the questions of what public leadership is conceptually and of how to measure it empirically.

Previous studies have paved the way to the measurement of public leadership (e.g., Fernandez 2005; Fernandez et al. 2010; Tummers and Knies 2016). Tummers and Knies (2016) were among the first to conceptualize leadership roles that are specific to the public sector and to develop a measurement instrument for public leadership. Building on a relationship-based view, they argue that leadership is reflected in the extent to which leaders support their followers in handling issues specific to the public sector. Accordingly, leadership in the context of the public sector is reflected in four roles: accountability, rule-following, political loyalty and network governance. After a thorough scale development procedure, Tummers and Knies (2016) suggest measuring each leadership role with four to six items as dimensions of the higher-order construct public leadership. The scale, in total consisting of 21 items, has proven satisfactory convergent, criterion-related and discriminant validity in a sample of Dutch public employees. The authors concluded with a call for more applications of the instrument in varied settings, which would allow for comparisons on the basis of a uniform and validated scale.

Public administration scholars often face practical difficulties, especially time and space restrictions, when adopting long measurement scales in a survey (van Engen 2017). Such restrictions in survey length and response time are particularly problematic in public leadership research for two reasons: First, public leadership theory is only at its beginnings and requires deeper explorations into the nomological network of public leadership, including its antecedents and consequences as well as relationships to other leadership approaches (Tummers and Knies 2016). Researchers are thus challenged to include many measures, often each with multiple items, in the same questionnaire. Second, the target groups of leadership questionnaires often suffer from lack of time. Given the shortage of personnel in many parts of the public sector, the workload of followers often does not allow for spending much time on completing extensive questionnaires. This applies even more to public officials and employees in leadership positions. These challenges might discourage researchers from adopting the public leadership scale or from measuring important correlates, which would impede progress in the scholarly field of public leadership. Researchers could get motivated to apply ad hoc measures of unknown validity.

The goal of the present study is, therefore, to develop a valid and reliable short form of the public leadership scale that can be applied in public administration research. To do so, we apply a 10-step scale reduction procedure (Stanton et al. 2002b) using data that we collected in two waves among public officials and employees in Germany ( $n = 2,291$  and  $1,726$ ). The result is a reliable and valid 11-item scale of public leadership.

## 2 | PUBLIC LEADERSHIP AND THE PUBLIC LEADERSHIP SCALE

Public administration scholars who call for more thorough explorations into public leadership (defined here as administrative leadership in public organizations) point to structural and behavioural differences between the public and the private sectors that might strongly affect the emergence and effectiveness of leadership (van Slyke and Alexander 2006; Orazi et al. 2013; van Wart 2013; Vogel and Masal 2015). Hence, we need to search for a distinct conceptualization and measurement of public leadership, 'rather than trying to retrofit existing concepts of leadership from business management or elective politics' (Getha-Taylor et al. 2011, p. i83). Tummers and Knies (2016) responded to this call by developing the public leadership construct and scale based on specific characteristics and values of the public sector and associated tasks of leaders. Compared to more heroic approaches in the past, their approach focuses on the leadership relationship with followers rather than on the traits of the leader. The resulting public

leadership scale measures the extent to which leaders encourage their followers to engage in behaviours that are crucial for public organizations to fulfil their tasks and to gain legitimacy. The construct comprises four roles, which are presented in the following sections.

## 2.1 | Accountability leadership

A major difference between public and private organizations is the accountability of public organizations towards a much broader range of stakeholders (van Slyke and Alexander 2006). Multiple stakeholder groups, such as citizens, voters, interest groups, courts, political bodies, and media, exert strong and often conflicting influence on public organizations. The concept of accountability increasingly extends beyond its core sense (of holding those in power accountable) and includes additional meanings (Bovens 2007). For example, it includes the responsiveness of public organizations to the needs of citizens, requiring public employees to respond to, explain and justify procedures and decisions (Mulgan 2000). Accordingly, Tummers and Knies (2016) consider accountability leadership as an important dimension of public leadership and define it as the extent to which leaders encourage their employees to justify and explain their actions to stakeholders.

## 2.2 | Rule-following leadership

Societal interests, once articulated in democratic procedures in the political system, translate into administrative agendas through rules, laws and governmental directives. 'Through rule design and implementation, public organizations distribute resources, empower employees, communicate fairness and trust, and impart understanding' (DeHart-Davis 2009, p. 909). Rule-following leadership is hence at the core of public leadership. Leaders who exercise rule-following leadership are those 'who encourage their employees to act in accordance with governmental rules and regulations' (Tummers and Knies 2016, p. 437).

## 2.3 | Political loyalty leadership

If the relationship between politicians and public employees is considered a principal-agent relationship, public employees are agents who are in charge of implementing the policies of their political principals. Principal-agent theory suggests that there are severe problems of incentives and control in such relationships due to information asymmetries (Jensen and Meckling 1976). Political loyalty is a possible remedy for agency problems and thus a precondition for an effective implementation of programmes. It means that public employees are willing to implement policies and directives even if they are not in line with their individual preferences and involve additional costs. Political loyalty leadership thus occurs when leaders 'encourage their employees to align their actions with the interests of politicians, even if this is costly for them' (Tummers and Knies 2016, p. 437).

## 2.4 | Network governance leadership

The fourth role reflects the recent trend towards collaborative forms of governance in and through networks. The growing complexity of societal problems requires public administration to collaborate with actors in business and civil society, rather than being the solitary actor in charge of solutions (e.g., Page et al. 2015). As a result, public managers increasingly find themselves in collaborative networks comprising stakeholders from different sectors and

backgrounds. Accordingly, public leadership includes the responsibility to facilitate such collaborations and to encourage 'employees to actively connect with stakeholders' (Tummers and Knies 2016, p. 437).

After an operationalization of these four public leadership roles, Tummers and Knies (2016) tested the psychometric properties of the scale using a sample of leaders and employees from various public sector organizations in the Netherlands. The initial pool consisted of 25 items, with four to seven items per dimension. After exploratory and confirmatory factor analyses, four items with poor quality were deleted. Convergent, criterion-related and discriminant validity was tested by correlating the four public leadership roles with other variables, including transformational leadership, perceived leadership effectiveness, organizational commitment, job satisfaction, work engagement, turnover intentions and organizational citizenship behaviour. The final result was a 21-item scale for the measurement of public leadership as exercised in four distinct roles.

Since its publication, the measurement instrument has resonated in the community of public leadership scholars in different ways and to different extents. While some scholars acknowledge the scale and suggest its application in future research (e.g., van Engen 2017; Crosby and Bryson 2018), others have selected single items from the scale to assess specific aspects of public leadership (e.g., Mathias 2017). As outlined above, the use of different versions of a scale without tested validity and reliability might impede scholarly progress. To the best of our knowledge, applications of the full scale are still rare (Mathias et al. 2019; Schwarz and Eva 2019), which might be due to the following reasons: on the one hand, leadership scholars might follow Ospina (2017) and prefer generic leadership constructs, because such constructs allow for a broader engagement with the field of general leadership studies, including comparisons between leadership in the public sector and other fields. On the other hand, and in more practical terms, the full scale might simply be too long to comply with the needs of empirical research in the public sector.

### 3 | OVERVIEW OF THE SCALE REDUCTION PROCEDURE

In public administration, survey instruments are often shortened ad hoc by individual researchers because validated scales are often overly long and time-consuming to respond to. Researchers frequently feel the need to shorten these scales to adjust to the limited space available in a questionnaire. This ad hoc process leads to a multiplicity of short versions of a scale, most of which are not adequately validated. This makes it difficult to determine whether a particular construct was measured adequately and whether the results based on these measures are valid. A good example of this problem is research on public service motivation (PSM) where various short versions of Perry's (1996) original instrument have been applied (Ritz et al. forthcoming).

To advise researchers on ways to shorten a validated scale while maintaining the validity of the measure, Stanton et al. (2002b) developed a comprehensive guide. We decided to use this procedure because it provides comprehensive step-by-step advice for researchers and at the same time is the most rigorous approach we found in the literature. Compared to approaches such as the one by DeVellis (2003), which was used in public administration research by van Engen (2017), it was developed to shorten an existing measure instead of developing a new one. It also emphasizes the importance of external and judgemental item quality in comparison to the internal item quality and includes a more elaborate statistical approach to compare the shortened and the full scales.

Stanton et al. (2002b) recommend a 10-step procedure to develop a short scale, which involves the collection of two datasets and several statistical techniques. The proposed procedure starts with the assessment of the external, internal and judgemental quality of each of the original items. External item quality focuses on the relationship between the items and other constructs. Internal item quality refers to the relationship between each item and the other items of the scale. Judgemental item quality refers to the subjective judgement of the items and takes the context of the scale and the environment in which it is administered into account. Stanton et al. (2002b) recommend using the assessment of the external, internal and judgemental item quality to select the items for the short version. We decided to aim for three items per leadership role to develop an instrument that allows for an appropriate assessment of the internal consistency of the measure (Liden et al. 2015; van Engen 2017, p. 517).

**TABLE 1** Scale reduction procedure

Step	Description	Procedure
1	Assessing the external item quality of the full scale	Correlation with job satisfaction and work engagement
2	Assessing the internal item quality of the full scale	Principal component analysis
3	Assessing the judgemental item quality of the full scale	Missing responses and researchers' judgement
4 & 5	Assessing the items of the short scale	Weighting of item quality results
6	Assessing the validity correlations of the short scale	Correlation with job satisfaction and work engagement
7	Assessing the internal consistency of the short scale	Principal component analysis
8 & 9	Rechecking the item-level performance with new data	Correlation with job satisfaction and turnover intentions; principal component analysis
10	Compare scale-level correlation matrices	Multi-group structural equation modelling

Note: Procedure based on the recommendations by Stanton et al. (2002b).

Table 1 gives an overview of the applied checks and the statistical or judgemental procedures used to develop and test a shortened version of Tummers and Knies' (2016) public leadership scale. All analyses were conducted using *R* version 3.6.2 (*R* Core Team 2019) and the *lavaan* package (Rosseel 2012).

## 4 | DATA

To test the original public leadership scale empirically, we conducted a survey in the district offices of a German city-state. The district offices primarily provide administrative services for residents and companies and are supervised by the city's ministries, which define rules and standards for the performance of delegated tasks. Interaction with and accountability to various stakeholder groups is thus particularly prevalent in the district offices' daily work. Citizens can participate in political decision-making through the election of district parliaments and various other forms of demand and control. Since the district offices provide citizens with a broad range of participation options in governmental programmes and decision-making, employees frequently find themselves in collaborative work settings. Altogether, we consider the district offices to be a fruitful empirical context for the study of all four dimensions of public leadership as introduced by Tummers and Knies (2016).

We collected our data using a paper-based survey among all government officials and employees of the district offices in two waves, with the first wave conducted from October to November 2017 ( $N = 7,381$ ) and the second wave from April to June 2018 ( $N = 7,651$ ). Due to a restrictive data privacy agreement with the participating organizations we are not able to publish the data used in this article. However, we published the analysis code along with synthetic data resembling the original data without violating participants' privacy (Nowok et al. 2016). Data and analysis code are available at <https://doi.org/10.17605/OSF.IO/RTV36>.

We defined leaders in terms of supervisory power, which implies that leadership, according to our understanding, is not limited to the very top of the district offices but is rather distributed across several hierarchical levels. In the first wave, 2,291 respondents completed the questionnaire, which corresponds to a response rate of 31.04 per cent. In the second wave, 1,726 respondents participated, corresponding to a response rate of 22.56 per cent. A total of 16.80 per cent of all participants were supervisors, 67.53 per cent were female, and 60.14 per cent were 46 years old or older. All sample characteristics are displayed in supplementary appendix S1. Since the data were collected in the context of a larger research project, not all constructs of the first wave were equally reflected in the

second wave. Supplementary appendix S2 displays all items used in this study together with their means and standard deviations. We also included the German translation of the items to support further use of the scale in German-speaking settings.

*Public leadership:* In the first wave, we measured public leadership with the initial 21-item scale developed by Tummers and Knies (2016). All items were translated from Dutch to German and back-translated by professional translators. After the pre-test and feedback from our field partner, we slightly adapted the wording of the items for political loyalty leadership. As part of the scale reduction procedure we shortened the scale to 11 items and tested the short version in the second wave. Details and results of the scale reduction process are presented in the following sections. Unless indicated otherwise, responses to all constructs were measured on a 5-point Likert scale from 1 = 'totally disagree' to 5 = 'totally agree'.

To assess the external validity of both the original and the short scale of public leadership, we measured the following additional constructs:

*Work engagement:* We measured work engagement of respondents relying on the German translation (Sautier et al. 2015) of the nine-item scale developed by Schaufeli et al. (2006). This scale was also used by Tummers and Knies (2016). Work engagement was measured in the first wave only.

*Job satisfaction:* Similar to Tummers and Knies (2016), we measured job satisfaction, but we applied the six-item scale developed by Tsui et al. (1992). In categories such as satisfaction with the supervisor or promotion opportunities, possible answers ranged from 1 = 'not satisfied at all' to 5 = 'very satisfied'. Job satisfaction was measured in both the first and the second wave.

*Turnover intentions:* Following Tummers and Knies (2016), we also asked respondents for their turnover intentions, but we used the three-item scale developed by Colarelli (1984). Turnover intentions were measured in the second wave only.

*Performance:* In both waves, we also measured the respondents' performance (as self-assessment) based on the seven-item scale developed by Sparrowe et al. (2001). We relied on this scale because it reflects the same categories (such as 'quality of work' or 'initiative') as the individual performance review used in the district offices.

## 5 | REPLICATION OF THE ORIGINAL SCALE

Before reporting the results of the scale reduction process, we demonstrate the comparability of the data collected in our first wave with the data of the original article by Tummers and Knies (2016). Table 2 compares the psychometric properties of the scale as measured in the first wave of our study with those reported by Tummers and Knies (2016). The table shows the internal consistency (Cronbach's alpha), the factor loadings of a confirmatory

**TABLE 2** Comparison of psychometric properties of the public leadership scale in the first wave and Tummers and Knies (2016)

	First wave (this study)				Tummers and Knies (2016)			
<i>Fit indices</i>	CFI = .990, TLI = .988, RMSEA = .112				CFI = .979, TLI = .975, RMSEA = .082			
<i>Leadership role</i>	<i>M</i>	<i>SD</i>	<i>Alpha</i>	<i>Loadings</i>	<i>M</i>	<i>SD</i>	<i>Alpha</i>	<i>Loadings</i>
Accountability	3.60	0.91	.911	.882	3.64	0.73	.934	.977
Rule-following	3.81	0.88	.834	.805	3.47	0.68	.821	.560
Political loyalty	3.17	0.93	.903	.551	3.07	0.71	.889	.395
Network governance	2.83	1.07	.949	.728	3.26	0.90	.958	.700

*Note:* *M* = Mean, *SD* = Standard deviation, *Alpha* = Cronbach's alpha, *Loadings* = Factor loadings of first-order factors on second-order public leadership factor in confirmatory factor analysis. Fit indices based on second-order factor analysis with diagonally weighted least squares (DLWS) estimator.

factor analysis and the model fit indices for both samples. The complete results of the factor analysis are presented in supplementary appendix S3.

The results indicate comparable properties with satisfying levels of internal consistency and overall fit. We also tested alternative factor models to determine whether a four-factor solution is most appropriate. The results reported in supplementary appendix S4 confirm that no alternative single-, two- or three-factor model is superior.

## 6 | SCALE REDUCTION

In accordance with the scale reduction procedure by Stanton et al. (2002b), the data collected in the first wave were used to assess the external, internal and judgemental item quality to select the items for the short version of the scale and to conduct initial tests on the short scale (steps 1 to 7). The short scale was then validated by using the data from the second wave (steps 8 to 10).

### 6.1 | Step 1: Assessing the external item quality

To assess the external item quality, we correlated each item of the full scale with participants' job satisfaction and work engagement. Tummers and Knies (2016) report correlations between the four public leadership roles and job satisfaction of  $r = .106$  to  $.272$  and between the four public leadership roles and work engagement between  $r = .055$  and  $.196$ . In the first wave, we find item-level correlations with job satisfaction between  $r = .119$  and  $.502$  and with work engagement between  $r = .133$  and  $.279$ . None of these correlations seem to be suspiciously low or high, and therefore items were not excluded at this stage. Supplementary appendix S5 shows the mean correlation between each item and job satisfaction as well as work engagement.

### 6.2 | Step 2: Assessing the internal item quality

To assess the internal item quality, we used principal component analysis (PCA). For each of the four public leadership roles, we calculated a separate PCA. Supplementary appendix S5 displays the resulting factor loadings of the items. To validate the results of the PCA, we used a graded response model analysis (Samejima 1969), which is based on item response theory. The resulting ranking of the items within their respective roles differed only marginally from the ranking based on the PCA. We therefore continued with the results of the PCA.

### 6.3 | Step 3: Assessing the judgemental item quality

To assess the judgemental item quality, we decided to inspect the proportion of missing responses per item because we assume that high proportions of missing responses indicate that items are not well understood by respondents. This, in turn, reduces the items' potential to measure the construct of interest. The data of the first wave show an average proportion of missing values of 3.14 per cent. While half of the items show a rate of 2.23 per cent or less, there are four items with rates of more than one standard deviation above the mean of all items (5.54 per cent, 6.63 per cent, 6.94 per cent and 9.60 per cent). All four items belong to the political loyalty leadership role. Given these relatively high non-response rates, we decided to exclude the three items for which most values were missing. We kept the item with a non-response rate of 5.54 per cent to ensure that political loyalty leadership is measured by at least two items.



## 6.4 | Steps 4 and 5: Selecting the items

Stanton et al. (2002b, p. 188) recommend sorting the items by the quality indices (step 4) and to 'use professional judgement to evaluate the items' quality scores and configure a suitable ... reduced length scale from among the top items' (step 5). Supplementary appendix S5 shows the raw values of these quality scores (mean correlation with job satisfaction and work engagement, PCA factor loading and proportion of missing responses) together with the ranking position within the respective public leadership role. The last column displays the mean ranking position over all three item quality scores and serves as the primary criterion for selecting the items. The bold items have been selected for the short version of the scale, which consists of three items per leadership role, except for political loyalty leadership, for which only two items have been selected due to the high non-response rates discussed above. We first tested the short version of the scale with data from the first wave and then with data from the second wave.

## 6.5 | Step 6: Assessing the validity correlations of the short scale

To assess whether the short version of the scale differs substantially from the full scale, we tested the correlation of the four public leadership roles with the measures of the same roles in the full scale as well as the correlations with job satisfaction and work engagement. To do so, we generated mean indices for each leadership role.

Table 3 reports the correlations of the measures of the four leadership roles based on the short scale with their respective counterparts in the full scale. The short and the full scales are strongly correlated with  $r = .916$  to  $.974$ , which indicates that they are highly related. The short and full versions of the scale also show comparable correlations with job satisfaction and work engagement. The two scales only differ substantially when the correlation between the short version of political loyalty leadership and job satisfaction ( $r = .268$ ) is compared with the respective correlation of the full scale ( $r = .219$ ). Since the two scales differ only marginally when the correlations with work engagement are compared ( $r = .175$  and  $.177$ ), we consider the difference to be negligible.

## 6.6 | Step 7: Assessing the internal consistency of the short scale

To assess the internal consistency of the short scale, we performed a PCA. We also checked the Cronbach's alpha values of the four public leadership roles. In addition to Stanton et al.'s (2002b) recommendations, we conducted a

**TABLE 3** Correlations between short and full scale as well as job satisfaction and work engagement based on data of the first wave

Leadership role	Corr (short, full)	Corr(short, job satisfaction)	Corr(full, job satisfaction)	Corr(short, engagement)	Corr(full, engagement)
Accountability leadership	.965***	.502***	.504***	.279***	.270***
Rule-following leadership	.969***	.363***	.365***	.201***	.213***
Political loyalty leadership	.916***	.268***	.219***	.175***	.177***
Network governance leadership	.974***	.402***	.418***	.274***	.283***

Note: Displayed correlations are Pearson's  $r$ . Short = short version of public leadership scale, full = full public leadership scale, engagement = work engagement; \* $p < .05$ ; \*\* $p < .01$ ; \*\*\* $p < .001$ .

**TABLE 4** PCA factor loadings, Cronbach's alpha, and CFA factor loadings of the short public leadership scale based on data of the first wave

Leadership role	PCA	Alpha	CFA
Accountability leadership	.671–.846	.841	.946
Rule-following leadership	.763–.826	.812	.726
Political loyalty leadership	.826–.884	.777	.622
Network governance leadership	.868–.906	.929	.698

Note: PCA = Range of factor loadings of the items on the respective leadership role in a principal component analysis; Alpha = Cronbach's alpha; CFA = Factor loadings of the leadership roles on a second-order public leadership factor in a CFA with DLWS estimator.

confirmatory factor analysis (CFA) and checked the model fit indices. The results of these tests are displayed in Table 4.

The results indicate good internal consistency of the short scale. In a PCA, the factor loadings range between .671 and .906 with none of the cross-loadings above .347 and only three loadings above .3. With values of .777 to .929, the Cronbach's alpha values of the four leadership roles are all well above the threshold of .7. The leadership roles also load with .622 to .946 on a common second-order public leadership factor, and the fit indices are acceptable (CFI = .996, TLI = .994, RMSEA = .084). The short version of the scale fits the data better than the full scale (CFI = .990, TLI = .988, RMSEA = .112) and the model reported by Tummers and Knies (2016) (CFI = .979, TLI = .975, RMSEA = .082).

## 6.7 | Steps 8 and 9: Rechecking item-level performance

After having successfully confirmed the validity and internal consistency with the original sample, we collected a second dataset (step 8) for which we only used the items of the short public leadership scale. The goal of this data collection was to verify the results of the previous tests. Following Stanton et al.'s (2002b) recommendations for the ninth step of the reduction procedure, we repeated the tests of steps 6 and 7 to assess the correlations with other constructs and the internal consistency with the data of the second wave.

The results of the correlation tests are shown in Table 5. Unfortunately, work engagement was not available in the data of the second wave, which is why we used turnover intentions, a construct also tested by Tummers and Knies (2016). There are only marginal differences between the first and second waves with regard to job satisfaction. Again, political loyalty leadership differs more than the other leadership roles. This might indicate that this role is more context-dependent than the other roles. Comparing the correlations between the short version and the results reported by Tummers and Knies (2016) does not yield any substantial differences, except for political loyalty leadership. Accordingly, Tummers and Knies' (2016) original hypothesis 2 (i.e., 'The four public leadership roles are positively related to organizational commitment, job satisfaction, work engagement and organizational citizenship behaviour and negatively related to turnover intentions'; p. 439) is broadly supported for job satisfaction, work engagement and turnover intentions.

Table 6 reports the tests for the internal consistency of the short public leadership scale based on the data of the second wave. Overall, the short scale shows good internal consistency with principal component factor loadings between .737 and .893, Cronbach's alpha between .739 and .924, and factor loadings in a CFA of .709 to .930. When comparing these values with those obtained in the first wave, we can even observe an increased internal consistency. The fit indices of the CFA are also slightly better than in the first wave (CFI = .997, TLI = .996, RMSEA = 0.071).

**TABLE 5** Correlations between the short version of public leadership and job satisfaction as well as turnover intentions based on data of the second wave

Leadership role	Job Satisfaction (second wave)	Job Satisfaction (first wave)	Job Satisfaction (Tummers and Knies 2016)	Turnover Intentions (second wave)	Turnover Intentions (Tummers and Knies 2016)
Accountability leadership	0.504***	0.502***	0.272**	-0.270***	-0.209**
Rule-following leadership	0.381***	0.363***	0.200**	-0.254***	-0.203**
Political loyalty leadership	0.389***	0.268***	0.106*	-0.184***	-0.057
Network governance leadership	0.442***	0.402***	0.238**	-0.229***	-0.103*

Note: \* $p < .05$ ; \*\* $p < .01$ ; \*\*\* $p < .001$ .

**TABLE 6** PCA factor loadings, Cronbach's alpha, and CFA factor loadings of the short public leadership scale based on data of the second wave

Leadership role	PCA	Alpha	CFA
Accountability leadership	.735-.813	.875	.930
Rule-following leadership	.812-.855	.860	.709
Political loyalty leadership	.737-.861	.739	.819
Network governance leadership	.817-.893	.924	.754

Note: PCA = Range of factor loadings of the items on the respective leadership role in a principal component analysis; Alpha = Cronbach's alpha; CFA = Factor loadings of the leadership roles on a second-order public leadership factor in a confirmatory factor analysis with DLWS estimator.

## 6.8 | Step 10: Comparing the scale-level correlation matrices

As a final step, Stanton et al. (2002b) recommend a multi-group structural equation model (SEM) to compare the scale-level correlation matrices between the full scale and the short version. Since they did not further specify the appropriate procedure, we followed the approach described by the authors in another article (Stanton et al. 2002a). We first randomly split the data of the first wave into two subsamples. Then we created mean indices for each of the four public leadership roles. For the first subsample, the mean indices are built from the full set of items. For the second subsample, we created the indices from the short scale. Accordingly, we built mean indices for the second wave, based on the short scale. This procedure enabled us to compare the full scale with the short scale in both the first and the second waves as well as the two short versions.

In a third step, we estimated multi-group SEMs comparing the full scale with the short scale in the first wave, comparing the full scale with the short scale in the second wave, and the short scale in the first and second waves. In the model, the four public leadership roles were treated as correlated independent variables affecting two correlated dependent variables. Since only a few potential outcome variables were included in both studies, we decided to use job satisfaction and self-reported performance as the dependent variables.

The estimation of the SEMs started with a baseline model, where all parameters, except for the correlation between job satisfaction and performance, were allowed to vary. In the next step, the internal correlations between the four leadership roles were fixed to be equal across groups. In the last model, the correlations between the leadership roles and the dependent variables were fixed.

**TABLE 7** Fit statistics and  $\chi^2$  difference tests for multi-group SEMs

Model	$\chi^2$	df	$\Delta \chi^2 (\Delta df)$	CFI	TLI	RMSEA
<i>Full scale in first wave vs. short scale in first wave</i>						
Baseline	0.988	1		1.000	1.000	.000
Internal correlations fixed	13.929	7	12.941 (6)*	.998	.992	.030
All correlations fixed	19.753	15	5.824 (8)	.999	.997	.017
<i>Full scale in first wave vs. short scale in second wave</i>						
Baseline	6.692	1		.999	.966	.064
Internal correlations fixed	60.407	7	53.715 (6)***	.989	.955	.074
All correlations fixed	68.497	15	8.090 (8)	.989	.979	.051
<i>Short scale in first wave vs. short scale in second wave</i>						
Baseline	2.162	1		1.000	.993	.029
Internal correlations fixed	32.863	7	30.701 (6)***	.995	.977	.052
All correlations fixed	46.324	15	13.461 (8)	.993	.987	.039

Note: df = degrees of freedom;  $\Delta \chi^2$  = difference of chi-square between the model and the less restricted model above;  $\Delta df$  = difference of degrees of freedom between the model and the less restricted model above; \* $p < .05$ ; \*\* $p < .01$ ; \*\*\* $p < .001$ .

To contrast the scale-level correlation matrices, we compared the baseline model with more restricted models. To assess whether the restricted models fit the data less—and consequently the correlation matrices differ—we used chi-square difference tests. The fit statistics of the multi-group SEMs and the difference tests are displayed in Table 7.

Table 7 shows that the restricted models of all three comparisons fit significantly less than the unrestricted baseline model. This indicates that the pattern of correlations between the leadership roles and the dependent variables job satisfaction and performance changed between samples. This is not unusual since the chi-square difference test is very sensitive, especially with such large samples as used in our study (Stanton et al. 2002a). To further assess the degree of fit of the restricted models, we also assessed global fit indices such as CFI, TFI and RMSEA. These indices show that—although the chi-square difference test is significant—the restricted models fit very well. Consequently, we regard the correlational patterns of the short version of the public leadership scale as adequate.

## 7 | DISCUSSION

The implications of this article are threefold. First, a shorter measurement scale for public leadership addresses a practical need in research as it makes it easier for researchers to assess the concept. Second, the application of the scale to a novel administrative context offers theoretical insights into the phenomenon of public leadership. Third, the implications also extend beyond research and theory as practitioners in public management also benefit from a shorter (and therefore more efficient) scale.

### 7.1 | Implications for research

As is often the case with nascent constructs, public leadership is plagued with measurement issues that might impede scholarly progress in this field. The scale developed by Tummers and Knies (2016) was a much needed step

forward in the study of public leadership. However, the length of the scale collides with practical limitations that public administration scholars often face when they design and implement questionnaires for their research. Although experimentation with different approaches to a construct has also virtues (McDaniel Sumpter et al. 2019), other subfields of public administration (e.g., research on PSM; Ritz et al. forthcoming) are cautionary examples of how researchers use short versions of measurement scales without proof of the validity and reliability of such ad hoc measures. To avoid a similar proliferation in public leadership studies, we applied a systematic scale reduction procedure and derived an 11-item version of the original public leadership scale (Table 8). This short scale retains the original four-dimensional structure of the construct and proves to be both valid and reliable.

Future research could investigate how to capitalize on these advantages even more. Since comparable research in the field of PSM shows that global measures are not generally inferior to multidimensional instruments (Wright et al. 2013), the development of unidimensional (and thus even shorter) measures of public leadership could be a potential avenue of research.

The results of our study also point to current problems with the public leadership scale. In particular, the political loyalty leadership dimension seems problematic as it exhibited by far the highest rate of non-responses. In a pre-test with practitioners, we received the feedback that these items read as if they reflected an inappropriate close relationship of administration with politics, which would violate the norm of impartiality and accountability to the public. For this reason, some participants feared that they implicitly committed a breach of civil service law when they scored high on these items. This might be a wording problem, which could reduce the importance of loyalty to individual decisions and decision-makers because it might evoke overly strong associations with personal favours done to politicians. Rather, commitment to the underlying democratic principles of decision-making in the politico-administrative system could be highlighted. More generally, this issue points to value conflicts in public leadership arising from the accountability of followers to various stakeholder groups at the same time. A promising avenue for further research could be the conceptualization of public leadership within the framework of paradoxical leadership (Lavine 2014; Zhang et al. 2015), which highlights the challenge for leaders to encourage the pursuit of several conflicting values at the same time.

**TABLE 8** Short public leadership scale

My supervisor ...

*Accountability leadership*

encourages me and my colleagues to explain our actions to various stakeholders.

encourages us to inform stakeholders of our way of working.

strives to ensure that we openly and honestly share the actions of our organizational unit with others.

*Rule-following leadership*

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

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

ensures that we accurately follow the rules and procedures.

*Political loyalty leadership*

encourages me and my colleagues not to jeopardize the relationship with political heads, even if that entails risks.

encourages me and my colleagues to defend political choices, even if we see shortcomings.

*Network governance leadership*

encourages me and my colleagues to invest substantial energy in the development of new contacts.

motivates me and my colleagues to regularly work together with people from our networks.

motivates me and my colleagues to develop many contacts with people outside our own department.

## 7.2 | Implications for theory

Beyond the practical advantages of the developed scale for researchers, our study provides further theoretical insights into the phenomenon of public leadership. To the best of our knowledge, we are the first to have measured public leadership in Germany. The results show the replicability of the scale and thus provide evidence that public leadership is to some extent generalizable across different administrative cultures. However, the levels of leadership behaviours in the four dimensions vary between the original and our sample. In particular, rule-following leadership is much stronger in Germany than in the Netherlands, whereas network governance leadership is much less pronounced. A possible explanation for these differences is the particularly strong *Rechtsstaat* tradition in Germany. In the rule-of-law culture, the main responsibility of public administration is the implementation of law (Kuhlmann and Wollmann 2019). On the one hand, strong hierarchy and legal orientation brings about obedience to formal rules (Reichard 2003), which facilitates rule-following leadership. On the other hand, the separation of state and society impedes collaboration across this divide (Aschhoff and Vogel 2019) and, as a likely consequence, network governance leadership. While some studies illustrate the emergence of collaborative forms of governance even in Germany (e.g., Royo et al. 2011; Parrado et al. 2013), public managers often engage in networking only for legitimization purposes or legal requirements (Royo et al. 2011). Our study thus reveals cross-country differences, which indicate that public leadership is embedded into, and intertwined with, the broader administrative structure and culture.

## 7.3 | Implications for practice

The advantages of a short scale for research practice, as outlined above, are also echoed in management practice. The shorter the scale, the more efficient its application for purposes of human resource management (HRM). As outlined by Tummers and Knies (2016), the scale is useful for HR professionals who want to assess at which levels leaders demonstrate the four public leadership roles. The scale can also be applied when the leadership talent of candidates for leadership positions is to be evaluated. Finally, the effectiveness of leadership development programmes can be measured by applying the scale before and after the training, or to participants and non-participants.

## 7.4 | Limitations

The following limitations of this study have to be acknowledged. First, the applied procedure of scale reduction required us to measure constructs other than public leadership (i.e., job satisfaction, work engagement, turnover intentions, performance). Although we used these constructs only for correlational (rather than causal) analyses, common method variance (CMV) might be an issue. We mitigated CMV to the largest possible extent by following the best practice recommendations by Podsakoff et al. (2012) in developing and proceeding with the questionnaire. It should also be noted that participants responded to our questionnaire in two different waves of the survey. However, since the same participants were invited to both waves, we could not completely rule out CMV resulting from a common source.

Second, the data used in this work were not collected exclusively for the development of a short version of the public leadership scale. As a consequence, we used correlates other than Tummers and Knies' (2016) to test the external validity of the scale, and for those correlates that were the same, we partly used different operationalizations. Therefore, differences in the correlations between public leadership and other constructs might be due to different operationalizations.

Third, we did not test the discriminant validity of public leadership against other leadership approaches. Since the goal of this study was the development of a short scale, we took the distinctiveness of public leadership for granted because Tummers and Knies (2016) established discriminant validity between public leadership and the most

widely used leadership construct (i.e., transformational leadership). Future research could test distinctiveness of public leadership from other leadership approaches.

Fourth, with data collected in local governments in Germany, this study also differs from the original research by Tummers and Knies (2016), who used data from a wide variety of public organizations in the Netherlands. However, despite the differences discussed above, the Netherlands and Germany still share many structural and cultural aspects in their administrative traditions (Painter and Peters 2010). We therefore call for broader use of the short public leadership scale in other contexts, for example, in cultures that strongly differ from Western European countries. In a similar vein, we encourage data collection in different types of public organizations.

## 8 | CONCLUSION

Despite these limitations, the development of a short public leadership scale fosters research on leadership in the public sector that has frequently been called for (van Wart 2013; Crosby and Bryson 2018). The short scale enables researchers to include multiple other constructs in the same questionnaire, which is necessary to expand on the nomological network of public leadership and to avoid overestimation due to omitted variable bias. It also facilitates measuring multiple leadership approaches and therefore allows researchers to study how public leadership relates to other approaches, such as servant leadership, authentic leadership or paradoxical leadership.

### ORCID

Dominik Vogel  <https://orcid.org/0000-0002-0145-7956>

Rick Vogel  <https://orcid.org/0000-0002-1361-5085>

### REFERENCES

- Aschhoff, N., & Vogel, R. (2019). Something old, something new, something borrowed: Explaining varieties of professionalism in citizen collaboration through identity theory. *Public Administration*, 97(3), 703–720.
- Bovens, M. (2007). Public accountability. In E. Ferlie, L. E. Lynn Jr., & C. Pollitt (Eds.), *The Oxford handbook of public management* (pp. 182–208). Oxford: Oxford University Press.
- Colarelli, S. M. (1984). Methods of communication and mediating processes in realistic job previews. *Journal of Applied Psychology*, 69(4), 633–642.
- Crosby, B. C., & Bryson, J. M. (2018). Why leadership of public leadership research matters: And what to do about it. *Public Management Review*, 20(9), 1265–1286.
- DeHart-Davis, L. (2009). Green tape and public employee rule abidance: Why organizational rule attributes matter. *Public Administration Review*, 69(5), 901–910.
- DeVellis, R. F. (2003). *Scale development: Theory and applications*. Thousand Oaks, CA: Sage.
- Fernandez, S. (2005). Developing and testing an integrative framework of public sector leadership: Evidence from the public education arena. *Journal of Public Administration Research and Theory*, 15(2), 197–217.
- Fernandez, S., Cho, Y. J., & Perry, J. L. (2010). Exploring the link between integrated leadership and public sector performance. *Leadership Quarterly*, 21(2), 308–323.
- Getha-Taylor, H., Holmes, M. H., Jacobson, W. S., Morse, R. S., & Sowa, J. E. (2011). Focusing the public leadership lens: Research propositions and questions in the Minnowbrook tradition. *Journal of Public Administration Research and Theory*, 21(1), i83–i97.
- Jensen, M., & Meckling, W. (1976). Theory of the firm: Managerial behavior, agency costs, and ownership structure. *Journal of Financial Economics*, 3(4), 305–360.
- Kuhlmann, S., & Wollmann, H. (2019). *Introduction to comparative public administration: Administrative systems and reforms in Europe*. Cheltenham and Northampton, MA: Edward Elgar.
- Lambright, W. H., & Quinn, M. M. (2011). Understanding leadership in public administration: The biographical approach. *Public Administration Review*, 71(5), 782–790.
- Lavine, M. (2014). Paradoxical leadership and the competing values framework. *Journal of Applied Behavioral Science*, 50(2), 189–205.
- Liden, R. C., Wayne, S. J., Meuser, J. D., Hu, J., Wu, J., & Liao, C. (2015). Servant leadership: Validation of a short form of the SL-28. *Leadership Quarterly*, 26(2), 254–269.

- Mathias, M. (2017). Public leadership in the United Arab Emirates: Towards a research agenda. *International Journal of Public Sector Management*, 30(2), 154–169.
- Mathias, M., Fargher, S., & Beynon, M. (2019). Exploring the link between integrated leadership-in-government and follower happiness: The case of Dubai. *International Review of Administrative Sciences*, 85(4), 780–798.
- McDaniel Sumpter, D., Greenberg, D., & Kim, S. (2019). The dark side of construct convergence: Navigating consensus, evolution, and practical relevance in theory building. *Academy of Management Perspectives*. <https://doi.org/10.5465/amp.2018.0153>
- Mulgan, R. (2000). 'Accountability': An ever-expanding concept? *Public Administration*, 78(3), 555–573.
- Nowok, B., Raab, G. M., & Dibben, C. (2016). synthpop: Bespoke creation of synthetic data in R. *Journal of Statistical Software*, 74(11), 1–26.
- Orazi, D. C., Turrini, A., & Valotti, G. (2013). Public sector leadership: New perspectives for research and practice. *International Review of Administrative Sciences*, 79(3), 486–504.
- Ospina, S. M. (2017). Collective leadership and context in public administration: Bridging public leadership research and leadership studies. *Public Administration Review*, 77(2), 275–287.
- Page, S. B., Stone, M. M., Bryson, J. M., & Crosby, B. C. (2015). Public value creation by cross-sector collaborations: A framework and challenges of assessment. *Public Administration*, 93(3), 715–732.
- Painter, M., & Peters, B. G. (2010). Administrative traditions in comparative perspective: Families, groups and hybrids. In M. Painter & B. G. Peters (Eds.), *Tradition and public administration* (pp. 19–30). London: Palgrave Macmillan. [https://doi.org/10.1057/9780230289635\\_2](https://doi.org/10.1057/9780230289635_2)
- Parrado, S., van Ryzin, G. G., Bovaird, T., & Löffler, E. (2013). Correlates of co-production: Evidence from a five-nation survey of citizens. *International Public Management Journal*, 16(1), 85–112.
- Perry, J. L. (1996). Measuring public service motivation: An assessment of construct reliability and validity. *Journal of Public Administration Research and Theory*, 6(1), 5–22.
- Perry, J. L. (2016). Is public administration vanishing? *Public Administration Review*, 76(2), 211–212.
- Podsakoff, P. M., MacKenzie, S. B., & Podsakoff, N. P. (2012). Sources of method bias in social science research and recommendations on how to control it. *Annual Review of Psychology*, 63, 539–569.
- Rainey, H. G., & Jung, C. S. (2015). A conceptual framework for analysis of goal ambiguity in public organizations. *Journal of Public Administration Research and Theory*, 25(1), 71–99.
- R Core Team (2019). *R: A language and environment for statistical computing*. Vienna, Austria. Retrieved from: <https://www.R-project.org/>
- Reichard, C. (2003). Local public management reforms in Germany. *Public Administration*, 81(2), 345–363.
- Ritz, A., Vandenabeele, W., & Vogel, D. (forthcoming). Public service motivation and individual job performance. In P. Leisink, L. B. Andersen, G. A. Brewer, C. B. Jacobsen, E. Knies, & W. Vandenabeele (Eds.), *Managing for public service performance: How people and values make a difference*. Oxford: Oxford University Press.
- Rosseel, Y. (2012). lavaan: An R package for structural equation modeling. *Journal of Statistical Software*, 48(2), 1–36.
- Royo, S., Yetano, A., & Acerete, B. (2011). Citizen participation in German and Spanish local governments: A comparative study. *International Journal of Public Administration*, 34(3), 139–150.
- Samejima, F. (1969). Estimation of latent ability using a response pattern of graded scores. *Psychometrika Monograph Supplement No. 17*. Richmond, VA: Psychometric Society.
- Schaufeli, W. B., Bakker, A. B., & Salanova, M. (2006). The measurement of work engagement with a short questionnaire: A cross-national study. *Educational and Psychological Measurement*, 66(4), 701–716.
- Schwarz, G., & Eva, N. (2019). The performance-enhancing effects of leadership style in Chinese public sector organizations. *Academy of Management Proceedings*, 2019(1), 18485. <https://doi.org/10.5465/AMBPP.2019.18485abstract>
- Sparrowe, R. C., Liden, R. C., Wayne, S. J., & Kraimer, M. L. (2001). Social networks and the performance of individuals and groups. *Academy of Management Journal*, 44(2), 316–325.
- Stanton, J. M., Sinar, E. F., Balzer, W. K., Julian, A. L., Thoresen, P., Aziz, S., ... Smith, P. C. (2002a). Development of a compact measure of job satisfaction: The abridged Job Descriptive Index. *Educational and Psychological Measurement*, 62(1), 173–191.
- Stanton, J. M., Sinar, E. F., Balzer, W. K., & Smith, P. C. (2002b). Issues and strategies for reducing the length of self-report scales. *Personnel Psychology*, 55(1), 167–194.
- Stazyk, E. C., & Davis, R. S. (2020). Transformational leaders: Bridging the gap between goal ambiguity and public value involvement. *Public Management Review*, 22(3), 364–385.
- Tsui, A. S., Egan, T. D., & O'Reilly, C. A. (1992). Being different: Relational demography and organizational attachment. *Administrative Science Quarterly*, 37(4), 549–579.
- Tummers, L. G., & Knies, E. (2016). Measuring public leadership: Developing scales for four key public leadership roles. *Public Administration*, 94(2), 433–451.
- Van Engen, N. A. M. (2017). A short measure of general policy alienation: Scale development using a 10-step procedure. *Public Administration*, 95(2), 512–526.



- Van Slyke, D. M., & Alexander, R. W. (2006). Public service leadership: Opportunities for clarity and coherence. *American Review of Public Administration*, 36(4), 362–374.
- Van Wart, M. (2013). Administrative leadership theory: A reassessment after 10 years. *Public Administration*, 91(3), 521–543.
- Vogel, R., & Masal, D. (2015). Public leadership: A review of the literature and framework for future research. *Public Management Review*, 17(8), 1165–1189.
- Wright, B. E., Christensen, R. K., & Pandey, S. K. (2013). Measuring public service motivation: Exploring the equivalence of existing global measures. *International Public Management Journal*, 16(2), 197–223.
- Yukl, G. A. (2006). *Leadership in organizations*. Upper Saddle River, NJ: Pearson/Prentice Hall.
- Zhang, Y., Waldman, D. A., Han, Y.-L., & Li, X.-B. (2015). Paradoxical leader behaviors in people management: Antecedents and consequences. *Academy of Management Journal*, 58(2), 538–566.

## SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of this article.

**How to cite this article:** Vogel D, Reuber A, Vogel R. Developing a short scale to assess public leadership. *Public Admin.* 2020;98:958–973. <https://doi.org/10.1111/padm.12665>