

# A validated measurement for felt relational accountability in the public sector: gauging the account holder's legitimacy and expertise

Sjors Overman, Thomas Schillemans & Stephan Grimmelikhuijsen

To cite this article: Sjors Overman, Thomas Schillemans & Stephan Grimmelikhuijsen (2020): A validated measurement for felt relational accountability in the public sector: gauging the account holder's legitimacy and expertise, Public Management Review, DOI: [10.1080/14719037.2020.1751254](https://doi.org/10.1080/14719037.2020.1751254)

To link to this article: <https://doi.org/10.1080/14719037.2020.1751254>



© 2020 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.



Published online: 06 May 2020.



Submit your article to this journal [↗](#)



Article views: 1222





View related articles [↗](#)



View Crossmark data [↗](#)

# A validated measurement for felt relational accountability in the public sector: gauging the account holder's legitimacy and expertise

Sjors Overman , Thomas Schillemans and Stephan Grimmelhuijsen 

School of Governance, Utrecht University, Utrecht, The Netherlands

## ABSTRACT

The effectiveness of formal public sector accountability mechanisms is largely predicated on the individual perception of accountability. In particular, the individual's experienced relationship to account holders is key in understanding the effects of formal accountability mechanisms. This article develops a measurement instrument for felt relational accountability in public administration. We measure perceived legitimacy and expertise of the account holder, as crucial relational dimensions applicable to various accountability relations. The instrument was tested and cross validated among two samples of Dutch public employees. We discuss theoretical implications of studying accountability at the actor-level and provide practical applications of the instrument.

**KEYWORDS** Felt accountability; accountability relations; scale development; legitimacy; expertise; behavioural public administration

## The issue: from formal to felt accountability

The study of accountability in public administration is extensive (Dubnick 2014). Public administration journals published, on average, over 150 articles on this topic each year from 2011 to 2017<sup>1</sup> (see also Kumar, Pandey, and Haldar 2020). This growing body of empirical studies on public accountability predominantly focuses on 'objective' and procedural aspects (Romzek, LeRoux, and Blackmar 2012). Many of these studies focus primarily on aspects such as legal obligations, the provision of information, or the use of audits, performance management and sanctions, which are all designed to affect (and control) the behaviours and decisions of policy actors (Brandsma 2014; Bovens, Schillemans, and Goodin 2014). As a result, our understanding of these formal and 'objective' dimensions has increased dramatically. For instance, we know how public organizations are subject to multiple accountability mechanisms and suffer from 'accountability overload' (Koppell 2005).

The debate in public administration has improved our understanding of accountability at the organizational level, but evidence from studies in the behavioural sciences shows that individuals perceive – and therefore, react to – identical accountability requirements in different ways (Frink and Klimoski 1998). Therefore, accountability 'functions in terms of a state of mind rather than a state of affairs' (Hall, Frink, and Ronald Buckley 2017,

**CONTACT** Sjors Overman  [s.p.overman@uu.nl](mailto:s.p.overman@uu.nl)

208). Indeed, the public administration debate has largely overlooked the fact that individuals may perceive and experience identical formal accountability settings in quite diverging ways (Tetlock 1992). This ties in with recent public administration studies documenting mismatches between formal accountability on the one hand and informal accountability on the other (Romzek, LeRoux, and Blackmar 2012; Page 2006). Only a handful of accountability studies in public administration implicitly use the actor-perspective (Romzek et al. 2014; Koppell 2005), while some others study individual leaders but overlook personal affectivity (Verhoest et al. 2010; Lupson and Partington 2011; Bach et al. 2016). Yet, individual accountability perceptions may have a strong impact, also in the public sector, for example in one-on-one interaction between police officers and citizens, or when agency CEOs make decisions that affect large organizations with significant consequences for policy. Further, accountability has been used in quantitative studies yet it has been measured in many different ways and such studies mostly intend to measure public accountability at the organizational, not individual, level (e.g. Brandsma 2014; Yang 2014).

This indicates an important gap in the public administration literature: individual civil servants have diverging perceptions of accountability and, thus, two individual civil servants may exhibit contradictory behaviour in response to the same formal accountability mechanism. Indeed, both psychological as well as public administration research documents that the anticipation of accountability has strong behavioural effects (Lerner and Tetlock 1999; Meijer 2000). Therefore, individual perceptions are crucial to our understanding of when and how formal accountability mechanisms 'work' (Bovens, Schillemans & 't Hart (2008); Dubnick & Frederickson (2011).

In order to study the individual's experience of accountability relations, this study develops a validated measurement instrument that reflects the relational nature of public accountability. Our instrument builds on existing instruments developed in psychological research. Scholars in psychology have studied individuals' 'felt accountability'<sup>2</sup>, and its antecedents and consequences (Hall, Frink, and Buckley 2017; Lerner and Tetlock 1999), as well as a measurement instrument (Hochwarter et al. 2005). Psychological research – and in particular, the measurements of felt accountability – has, however, focused on the anticipation of a future accountability moment only, without reflecting the relational nature of accountability (Ebrahim 2003). Therefore, such measures alone have limited value for measuring felt accountability in the public sector, since many public administration studies have demonstrated, accountability is quintessentially a *relational* concept, and it matters a great deal *to whom* an actor is accountable (Bovens 2007). Therefore, one important contribution of the current study is the development and validation of an instrument that is able to measure the experience of the accountability relationship and to compare these measurements between individuals.

The accountability relationship, in general, has been conceptualized as the relationship between an account holder or principal and the account giver or agent (the actor) (Strøm 2000; Bovens 2007). In public administration studies, numerous scholars have pointed out the multiplicity of those relationships. Policy actors operate in 'webs' of accountability (Page 2006), are confronted with accountability claims from above, aside, and below (Bovens, Goodin, & Schillemans 2014, Hooghe & Marks 2003), and adhere to political, legal, bureaucratic and professional accountability standards (Romzek and Dubnick 1987).

In a public sector context, the accountability relationship generally is different from other hierarchical relationships in organizations, such as between employer and supervisor (cf. Waldo 1948). On the one hand, civil servants are accountable to democratically elected political principals. On the other hand, civil servants are often (quasi-) autonomous and only accountable to external account holders, such as in the case of public service delegation to an executive agency (Curtin 2007; Verhoest et al. 2010). Therefore, we need to detail the relational aspect of accountability in the conceptualization and measurement of felt accountability in the public sector and acknowledge that actors can relate quite differently to the various account holders. The relational dimension needs to be added to our measurements.

In individual's assessment an accountability relationship, the key issue is the perception of the account holders' *ability to hold the actor accountable* (Mulgan 2003). This perception relies on two crucial dimensions of public sector accountability relations (cf. Bovens, Schillemans, and Hart 2008): (1) the perceived *legitimacy* to exercise authority, and (2) the perceived capacity based on *expertise* to do so, effectively. The ability to monitor and evaluate strikes the heart of the principal-agent problem and constitutes the premise of most accountability systems (Strøm 2000; Gailmard 2014). It is, thus, important to investigate whether actors recognize the legitimacy or 'moral authority' (Mulgan 2003) of the account holder scrutinizing its behaviour. Furthermore, the account holder needs to be identified as possessing some degree of expertise to be able to oversee and evaluate the task delegated to the agent. The role of technical expertise in the implementation of policy is significant, and so is the deference to expertise in public accountability relations (Romzek and Dubnick 1987, 229–30).

The two relational dimensions of legitimacy and expertise serve as expansions to the existing concept and measurement of felt accountability introduced from psychology. In the next sections, we define and develop the concept of felt relational accountability and highlight its importance for the current accountability debate. Next, we discuss the psychological concept of felt accountability and subsequently expand this by developing legitimacy and expertise as important additional features. We, then, develop a set of eight items to measure the concept of felt relational accountability, to be used alongside the existing Hochwarter et al. (2007) scale. These indicators have been tested and validated in two separate sub-samples in the Dutch public sector, with a total N of 2,702, covering respondents from a wide variety of public organizations – such as the police and municipalities – and functions – such as managers and professionals. We conclude by identifying potential applications of the scale and avenues for further improvement.

### **The challenge: felt relational accountability in public administration**

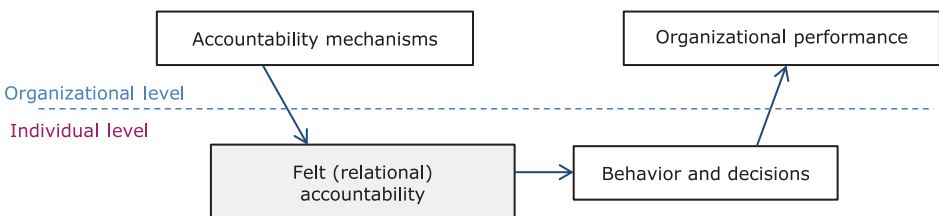
The definition of accountability is, as most other concepts in the social sciences, complex and contested. Over the years, however, a degree of conceptual consensus is emerging (Bovens, Schillemans, and Goodin 2014). Many current studies in public administration refer back to Bovens' definition (2007, 450–52) which, in turn, is strongly comparable to some other leading conceptions of accountability (Romzek and Dubnick 1987; Mulgan 2000; Olsen 2013). Bovens states that accountability is characterized by an actor and a forum, where the actor has an obligation to provide information and clarification about its conduct, the forum can ask questions, pass judgement and make the actor face consequences of its behaviour via the use of sanctions.

Governments have devised various accountability *mechanisms* aiming to nudge, push or force actors in government towards accountable behaviour. Many recent studies have accordingly mapped or evaluated the landscape of formal accountability mechanisms surrounding various administrative organizations (Verhoest et al. 2010; Koop 2014; Bovens, Schillemans, and Goodin 2014; part V). These include agencies providing information to ministries (Hong, Kim, and Son 2019; Koop 2011), information availability on websites (Wong and Welch 2004), or site visits by evaluators (Carman 2008), to name just a few examples. These formal mechanisms may be an end in themselves as they are constitutive of democratic governance. Even more commonly, they are treated as means to various ‘promises’ of accountable governance (Dubnick and Frederickson 2011). Accountability mechanisms, then, are independent variables potentially leading to various, normatively desirable, dependent variables.

Formal accountability mechanisms, however, remain utterly symbolic if they do not lead to real accountability *processes*, in which actors actually explain their behaviour. Romzek, LeRoux, and Blackmar (2012, 2014) show that informal accountability is highly relevant in the public sector. ‘Individual actions and behaviors [reflecting] informal accountability play significant complementary roles to formal accountability’ (Romzek, LeRoux, and Blackmar 2012, 443). And formal accountability arrangements align with expectations of various external stakeholders, which changes the ‘politics of public management’ and may result in major differences on the ground between formally comparable cases (Page 2006, 193–4). Important social and informal accountability processes affect individual behaviours, yet these processes escape our attention when we focus exclusively on formal accountability mechanisms.

We contend that the causal chain linking observable formal accountability mechanisms to desirable (= accountable) organizational outputs and outcomes entails a number of intermediate steps. The design of accountability mechanisms for the various organizations in the public sector is predicated on the guiding anticipation that 1) *de jure* accountability mechanisms trigger 2) *de facto* accountability processes, which in turn, 3) affect individual *perceptions* of accountability, which 4) have an impact on decisions and behaviours by those individuals in organizations, ultimately, 5) ‘producing’ organizational outputs, outcomes and performance, which, then, can be understood to be desirable (or accountable), or not. Put differently, readily observable accountability mechanisms should ultimately contribute to observable outcomes in the public sector via relational intra-organizational processes and individual perceptions of those accountability relations.

The causal link between formal and informal accountability mechanisms on the one hand and desirable outcomes on the other thus entails various intermediary steps in



**Figure 1.** Coleman's boat applied to accountability.

which the individual's perception of accountability (felt relational accountability) is the crucial linking pin. This is visualized in [Figure 1](#).

[Figure 1](#) is inspired by Coleman's (1990) image of the 'boat' linking macro-level conditions and outcomes via micro-level conditions and outcomes. This analytical frame helps to clearly distinguish between organizational (or institutional) and individual levels of analysis and also illuminates how these are connected. Above all, this approach helps to understand how collective, macro-level phenomena can be studied by focusing on individual behaviours and perceptions, as is common in the growing body of studies in behavioural public administration (Grimmelikhuijsen et al. 2017).

In [Figure 1](#), the organizational level displays formal accountability mechanisms for organizations and the desirable organizational outcomes or 'promises' of accountability (Dubnick and George Frederickson 2011). The individual level features the felt relational accountability: the extent to which individual actors perceive their institutional accountability environment to be salient, and ensuing individual behaviours which collectively add up to organizational outcomes. In other words, *felt relational accountability* is key in our understanding of the causal link between institutional accountability arrangements on the one hand, and behavioural outcomes on individual and aggregated levels in public administration on the other.

### The base: felt accountability

Felt accountability in psychology refers to the belief held by actors that they will be held accountable. Hall and Ferris (2011, 134) define it as 'an implicit or explicit expectation that one's decisions or actions will be subject to evaluation by some salient audience(s) with the belief that there exists the potential for one to receive either rewards or sanctions based on this expected evaluation.' These beliefs include the anticipation of a future instance of accountability to some salient account holder, as well as the beliefs that an actor holds about the account holder and its ability to hold the actor to account. Some studies in public administration acknowledge this or closely related concepts (Klingner, Nalbandian, and Romzek 2002; Lupson and Partington 2011).

The systematic exploration and measurement of felt accountability has advanced in social and organizational psychology (Aleksavska, Schillemans, and Grimmelikhuijsen 2019). The concept of felt accountability has numerous strong effects on decisions and behaviours in experimental studies (Frink and Klimoski 1998; Hall, Frink, and Buckley 2017). In their recent review of experimental accountability research, Hall, Frink, and Ronald Buckley (2017) focus exclusively on felt accountability as the upshot of this productive branch of the academic literature (see also Patil, Vieider, and Tetlock 2014; Harari and Rudolph 2017). Felt accountability can either denote a general state-of-affairs, in which the actor expects constant accountability for all actions and decisions, a situational state in which a specific action or decision is likely to be scrutinized, or a routinized state in which actors always expect accountability for specific actions. This expectation of accountability includes the expectation of the account holder's judgement, as well as potential use of sanctions.

In experimental studies, felt accountability has been measured in various ways, but Hall and colleagues argue that (2017, 214): 'in general, [the measures] do not reflect the evolving, social, and relationship-based nature of accountability within organizations.' In some studies, felt accountability is not measured directly but used as a manipulation in the design of an experiment. Some participants in an experiment are, for instance,

told that they will have to explain their behaviour to someone after the experiment, while some other participants are not made to expect accountability (Peng, Dunn, and Conlon 2015). Other studies do not measure felt accountability directly but, rather, measure prerequisites for accountability. Respondents are, for example, asked to indicate whether their work is visible to their account holder (usually a direct hierarchical supervisor), whether mistakes can be attributed to individuals and whether mistakes are punished (Mero, Guidice, and Werner 2014).

The most commonly used measure is based on an instrument developed by Hochwarter et al. (2007).<sup>3</sup> This scale is useful and has often been used effectively, yet, as Hall, Frink, and Ronald Buckley (2017) conclude, The Hochwarter scale is empirically unidimensional and is primarily about the *anticipation of accountability*, denoting whether actors expect that they will (or are likely to) be held accountable for their decisions or actions. This expectation of accountability includes the expectation of the account holders' and potential sanctions. The anticipation of future accountability is also relevant in a public sector context (cf. Friedrich 1963 and his *Law of anticipated reactions*, empirically corroborated by Ford and Zelditch 1988; Meijer 2000). At the same time, the environment in which an actor perceives future accountability also involves beliefs about additional aspects, most prominently, the beliefs an actor holds about the account holder to whom (s)he is accountable.

### **The extension: felt *relational* accountability**

As discussed in the introduction, accountability relationships in the public sector are generally peculiar and not directly comparable to 'ideal-typical' relationships between employers and supervisors (cf. Waldo 1948). The public accountability environment is characterized by a democratic context (*ibid.*), a multiplicity of accountability mechanisms and account holders (Romzek and Dubnick 1987; Koppell 2005; Bovens, Goodin, & Schillemans 2014), and an extra-organizational nature of many accountability mechanisms. As noted by Hall, Frink, and Ronald Buckley (2017), the Hochwarter scale refers to the top management as account holder only. This is, thus, insufficient for public sector settings and calls for flexibility in the measurement instrument regarding the formulation of the questions about the relational part of felt accountability. Studies also suggest that account holders can differ widely in their ability to hold agents accountable (Mulgan 2003).

Moreover, psychological and public administration research acknowledge the inherent relational aspect of accountability. Most definitions of accountability in public administration stress the relational nature (Romzek and Dubnick 1987; Mulgan 2000; Bovens 2007; Olsen 2013). Likewise, the most common definition of felt accountability in psychology, on which the Hochwarter scale has been based, stipulates that felt accountability is: 'The 1) *implicit or explicit expectation* that one's decisions or actions will be subject to evaluation by 2) *some salient audience(s)* with the belief that there exists the potential for one to receive either rewards or sanctions based on this expected evaluation.' (Hochwarter et al. 2007: 227, emphasis and numbers added). The literature in public administration and the behavioural sciences, thus, clearly agree on a definition of accountability consisting of two parts: the *expectation* of future accountability and the salience of, in our words, the account holder.

The existing measurement in psychology almost exclusively focuses on the first part of the definition: the anticipation of future accountability. The only slight reference to

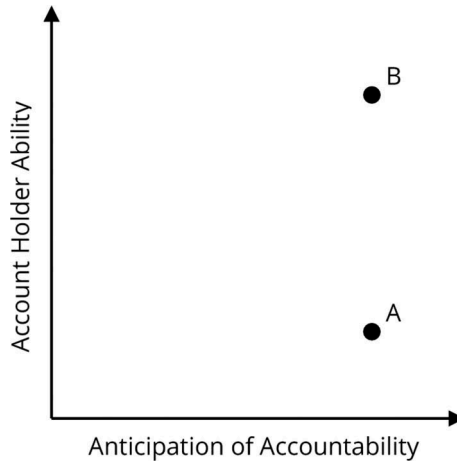
the second part of the definition, the salience of the account holder, is that 2 of the 9 items mention 'top management' in items 2 and 3 (see note 2). The salience of top management is then implied; and not measured. This also makes it hard to use this scale in other accountability settings in the public sector where strict hierarchy may be absent. This limitation of the measurement scale has been noted in the review of felt accountability in the behavioural sciences by Hall et al. (2017, 241) claiming that the Hochwarter scale is unidimensional and ignores the nature of the audience. That is, the current challenge is not only to measure the *expectation* of accountability, but also the *salience* of the account holder.

We, therefore, propose to expand the measurement of felt accountability beyond the actor's *anticipation* of accountability with the actor's perception of the *ability* of a given account holder to hold the actor accountable (Mulgan 2003). This adds the salience of the account holder which is implied in Hochwarter's definition yet is absent from the original measurement. This perception of the account holders' ability relies on how the account holder is able to nurture the impression of both the 'authority of ideas' and the 'authority of sanctions' (Simon 1997, 187). Two core elements in the actor's perception of the forum's ability to hold the actor to account (cf. Bovens, Schillemans, and Hart 2008) are (1) the perceived *legitimacy* to exercise authority and (2) the perceived capacity based on *expertise* to do so effectively. Both concepts are central to Weber's classic conception of bureaucracy (Miller 1970).

The actor's perception of the account holders' legitimacy or (Mulgan 2003, 10–11) is the first feature of felt relational accountability. This perspective on moral authority owes to Carpenter and Krause (2015) conceptualization of transactional authority. That is, an account holders' authority can only influence the actor's behaviour based on the actor's 'sanctioned acceptance' (Simon 1997) of the authority. Without the (implicit) acceptance of authority, the actor will not be influenced by the account holder. In line with their approach, our point of departure to conceptualize authority lies not in the formal sense of authority, but in the perceptual sense of legitimate authority. Following Suchman's (1995) classic definition, we understand legitimacy to be a 'generalized perception or assumption', held by the accountable actor, that the account holders' scrutinizing role is 'desirable, proper or appropriate' and in line with the 'norms' and 'values' of the democratic system of government. Legitimate authority should be seen as a 'contingent grant' (Presthus 1960, 88), that requires input based on respect of the actor and, therefore, bears the actor's 'willingness to obey' (Simon 1997, 180). In other words, the actor's evaluation of the legitimacy of an account holder relies on the relationship that is built and nourished between both parties and with the consistency of 'expectations' (Dubnick and Romzek 1993) institutionalized in their relationship. Experimental research in social psychology demonstrates that actors respond in a positive way to account holders they perceive as legitimate (Tyler 1997; Lerner and Tetlock 1999). This is of crucial relevance to public administration accountability relationships which are often shifting, contested and ambiguous (Olsen 2017).

The second feature is the *perceived expertise* of the account holder. Its capacity to oversee the agent and to prevent the agent from 'shirking or sabotage' (Brehm and Gates 1999) is a necessary precondition for any accountability relationship to work (Strøm 2000), and there is ample evidence of failing account holders in public administration (Schillemans and Busuioc 2015). The account holder needs to possess some degree of expertise to be able to oversee and evaluate the task delegated to the agent. Empirical research has demonstrated that attributing greater expertise to an account holder has a positive effect on the cognitive effort an actor expends in response to accountability (Schillemans 2016). In other words, an actor considering the account





**Figure 2.** Dimensions of felt accountability: anticipation and perceived ability of the account holder.

Note: Point A – anticipated accountability to a non-salient forum; Point B – anticipated accountability to a salient forum.

holder to be an expert is a key part of felt relational accountability. The role of technical expertise in the implementation of policy is significant, and so is the deference to expertise in public accountability relations (Romzek and Dubnick 1987, 229–30).

For public bureaucracies as ‘professional organizations’ (Scott 1965), the question whether the organization perceives that the account holder fully understands the nature of their tasks may strongly affect how they respond to claims for accountability. Examples of bureaucracies criticizing anticipated and legitimate calls for accountability for insufficient expertise are for instance public managers criticizing their political superiors for ‘paper[ing] over . . . differences by issuing superficially attractive but incoherent and incomplete policy directives’ (Donahue and Moore 2012; quoted from Alford et al. 2017), or a governmental climate scientist having to come to terms with ‘facts and politics’ (Lowe 2007, 63).

We, thus, develop a measurement instrument in addition to the unidimensional Hochwarter-scale, which only measures the anticipation of accountability. See Figure 2 for a graphic representation of the model. We add a valid measurement of the experienced legitimacy and expertise of the account holder. We also include more flexibility to replace the account holder in the questions included in the measurement. We, then, test this instrument in various public sector settings to demonstrate its added value.

### **The goal: a reliable scale for felt relational accountability in public administration**

In public administration studies, ‘accountability’ has been measured in many ways. Several scholars claim that the measurement of accountability is underdeveloped in public administration (Dubnick 2014; Klingner, Nalbandian, and Romzek 2002). Brandsma (2014) provides an overview of existing quantitative studies on accountability. Some public administration studies measure specific formal dimensions of accountability (Koop 2014; Verhoest et al. 2010; Overman, Van Genugten, and Van Thiel. 2015; Bovens and Wille 2016). Other studies apply a more sociological empirical

perspective and analyse accountability processes as they evolve (Schillemans 2011; Romzek, LeRoux, and Blackmar 2012).

Yet, in general, accountability studies have been based on single cases and use specific and situationally appropriate yet hard to compare, let alone aggregate, measurements (Yang 2012). Against this background we believe there is good reason to develop a scale that can be applied in different quantitative accountability studies. Such a measurement facilitates the comparison of diverging perceptions of accountability mechanisms (Tetlock, Skitka, and Boettger 1989), and allows for a clear analysis of antecedents and effects of that perception (Lerner and Tetlock 1999).

### ***Item development***

We developed and validated our measurement instrument using commonly used recommendations in public administration research (Hinkin 1998; DeVellis 2009). The development of validated scales more generally could improve the quality of public administration research (Pandey and Scott 2002). Indeed, recently various scales have been developed for common public administration themes, such as perceived public service motivation (Kim et al. 2013), bureaucratic reputation (Overman, Busuioc, and Wood 2020), and perceived red tape (van Loon et al. 2016). Such validated scales facilitate the comparison of phenomena, such as felt relational accountability in our case, between organizations, policy sectors, and countries.

After a thorough definition of the concept and the identification of its constituting dimensions, we developed survey items to measure the various dimensions. We started by generating a list of items. These items were assessed by three external experts who have specialized in accountability research. The items were subsequently presented and discussed in two additional group sessions during international conferences with multiple scholarly expert participants.

Ten items were developed to capture account holder expertise and legitimacy, based on a review of accountability research in behavioural studies and public administration (Schillemans 2016). In order to prevent socially desirable answers, we do not directly ask the assessment of the account holders' legitimacy and expertise. Rather, we focus on indirect measures and consequences. The items were answered on a 5-point Likert scale, see Table 1 for the items and descriptive statistics of the responses. The survey questions are formulated in a responsive way, such that the account holder to which an actor is accountable can be adapted. The adaptable phrases are formulated between brackets (see Table 1). 'The account holder' can be replaced, for example, by 'an official in the ministry of Education', or 'the Court of Audit'. We briefed the respondent in an introductory screen as follows:

'Please consider the important tasks in your current job. We will now ask you some questions about your work – that is, the tasks that you consider important – and your contact with [the account holder]. When we ask about your work, you may also refer to the work you do in a team or group in your answer.'

### ***Samples***

A questionnaire containing the items was administered to a total of two samples of Dutch public sector employees. The questionnaire, furthermore, included the original Hochwarter-scale of felt accountability, which was established as a reliable scale with

**Table 1.** Initial item pool.

			Sample 1 (calibration)	Sample 2 (validation)
Legitimacy	L1	When [the account holder] changes their views, I have to comply with the new reality.	3.75 (0.86)	3.83 (0.87)
	L2	It is a good thing, that I am ultimately (also) accountable to [the account holder].	3.90 (0.81)	3.97 (0.81)
	L3	I am willing to work in the interest of [the account holder].	3.79 (0.91)	3.89 (0.86)
	L4	When [the account holder] has an opinion in matters relating to my work, I consider that opinion important.	2.72 (1.09)	2.62 (1.11)
Expertise	E1	[The account holder] has sufficient substantive or technical expertise to oversee/evaluate my work.	2.52 (0.99)	2.46 (1.03)
	E2*	I have to simplify information for [the account holder] to understand it.	3.53 (0.97)	3.52 (0.99)
	E3	[The account holder] provides constructive feedback about my work.	2.86 (0.90)	2.85 (0.92)
	E4*	The opinions of [the account holder] about my work regard core issues, rather than trivial issues.	3.23 (0.99)	3.40 (1.00)
	E5	Opinions of [The account holder] are generally unambiguous.	2.69 (0.89)	2.54 (0.92)
	E6	Faced with a dilemma, I can ask [the account holder] for advice.	3.39 (0.98)	3.46 (0.98)

Note: item means; standard deviations in parentheses. \*E2 and E4 were dropped in the final instrument, see below.

Cronbach's  $\alpha = .73$  for both samples combined.<sup>2</sup> The first study was conducted among public sector employees effectively working below the national level in local and regional governments (municipalities and provinces) as well as in the police in November 2016. We fielded a second wave among employees in the national government (ministries and executive agencies) in February 2017. The first wave contains 1,983 responses (response rate: 36.3%), and the second wave contains 1256 responses (response rate: 35.3%). *Internetspiegel*, a Dutch government-owned personnel research programme, took the samples from an existing panel of Dutch respondents working in the public sector and provided us with the data.

For age and salary: mean scores, standard deviations in parentheses.

We categorized the subsamples according to the organization for which respondents work, as the organization determines the external account holder to whom employees render account. Municipal civil servants are formally accountable to the mayor and aldermen; provincial civil servants to the Provincial Executive; police officers to the minister of Security and Justice, ministerial civil servants to the government; and executive agency employees to their parent departments. These were the account holders that were automatically inserted in the survey questions for those employees who reported to work outside of the support staff and internal operations. For the latter respondents, we measured internal accountability, instead of external accountability; as those respondents are unlikely to be in contact with the external account holder. The account holder inserted in their survey questions was 'your supervisor'. More detailed sample characteristics can be found in [Table 2](#).

### Calibration

The measurement instrument was developed in a two-step strategy. First, we calibrated the initial model to the data from the first sample. Calibration involves the adaptation

**Table 2.** Sample characteristics.

Organizations	Sample 1 (N = 1,977)	Sample 2 (N = 1,256)
	Municipalities, provinces, police	Ministries, executive agencies
Age	53.00 (8.47)	52.29 (8.10)
Monthly salary (k€)	3.77 (1.13)	4.29 (1.27)
Female	33%	27%
Education		
– Lower	6%	4%
– Middle	29%	19%
– Higher	65%	77%
Supervisor (respondent is supervising one or more employees)	17%	20%
Operational management or support staff	20%	11%

For age and salary: mean scores, standard deviations in parentheses.

of the initial model based on the data in the first sample (Kline 2015). We use the second sample to test and cross-validate the model to prevent overfitting or capitalizing on chance based on the available data in the first sample (MacCallum, Roznowski, and Necowitz 1992).

For Sample 1 we retained only respondents with an external account holder. That is, we removed respondents that were not dealing with external accountability: public servants not working as operational management and support staff ( $n = 1590$ ). We used the lavaan package in R for confirmatory factor analysis to test our models (Rosseel 2012). First, we evaluated item fit using the communality ( $R^2$ ) and model fit using absolute fit indices (RMSEA and SRMR), as well as incremental fit indices (TLI and CFI). A RMSEA value lower than .06 and an upper bound of the 90% confidence interval of .08 are generally considered indicative of good fit, as are SRMR indices of .05 and lower (Hooper, Coughlan, and Mullen 2008). For the incremental fit indices, values above .90 indicate moderate fit, and values above .95 indicate good model fit (Hu and Bentler. 1998).

The initial model, which included all items listed in Table 1 – had a moderate fit with the data: RMSEA = .058 (90% CI [.051-.066]); SRMR = .042; TLI = .917; CFI = .937. Therefore, we calibrated the model to better fit the structure of the data. In our calibration, items E2 (simplify information) and E4 (opinions on core issues) were dropped from the initial model due to low communality – the proportion explained variance in the items by the factor:  $R^2 = .01$  and  $.05$ . In the calibrated model, we retain the four remaining items for each dimension (the legitimacy dimension had only 4 items). The calibrated eight-item model has a good fit: RMSEA = .053 (90% CI [.043-.063]); SRMR = .032; TLI = .954; CFI = .969. The second model also has increased incremental fit indices compared to the initial model, which emphasizes the model improvement. These findings were underscored by the differences in BIC (model 1: 38,290.54, model 2: 30,042.34). See Table 3 for the factor loadings.

## Validation

To validate the measurement equivalence of the calibrated model, we acquired fit indices from data from the second sample. In this sample, we also retained only the respondents with an external account holder ( $n = 1,112$ ). Again, the eight-item model had a good fit with the validation sample: RMSEA = .051 (90% CI [.039-.063]);

**Table 3.** Standardized factor loadings.

Dimension		Sample 1 (calibration)	Sample 2 (validation)
Legitimacy	L1	.44	.44
	L2	.57	.54
	L3	.67	.63
	L4	.74	.73
Expertise	E1	.56	.59
	E3	.78	.67
	E5	.60	.65
Model fit	E6	.60	.62
	SRMR	.032	.031
	RMSEA [90% CI]	.053 [.043-.063]	.051 [.039-.063]
	TLI	.954	.953
	CFI	.969	.968

SRMR = .031; TLI = .953; CFI = .968. The fit of the model in all of the subsamples is a good indication that the measurement does not need to be modified further, and that the instrument is able to measure felt relational accountability in various contexts within the public sector. Table 3 presents the standardized factor loadings.

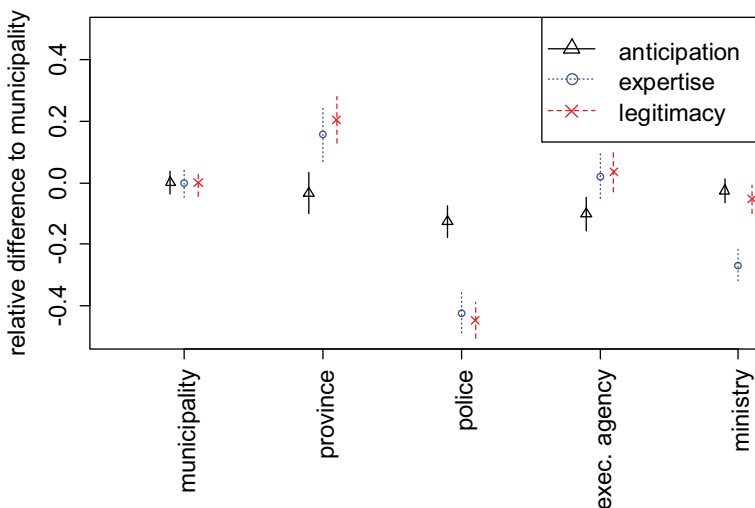
We tested reliability and convergent validity by examining factor loadings, Cronbach's alpha, and coefficient omega. Factor loadings of .3 are generally accepted as significant, and loadings of .5 and higher are considered practically useful. The standardized factor loadings varied between .44 and .78, with the majority greater than .5, see Table 3. We evaluated factor reliability using Cronbach's  $\alpha$  for the combined samples. Both factors were reliable ( $\alpha_{\text{expertise}} = .73$ ,  $\alpha_{\text{legitimacy}} = .69$ ). The composite reliability of the model was further evaluated using coefficient  $\omega$  (Raykov 2001), as this measure does not assume tau-equivalence – an equal loading of all items on the same factor – in contrast with Cronbach's  $\alpha$ . The values for coefficient  $\omega$  were acceptable, as well ( $\omega_{\text{expertise}} = .73$ ,  $\omega_{\text{legitimacy}} = .70$ ).

To test discriminant validity of the model, we compared the two-factor model to a single factor and a second order model that used a single latent factor on which account holder legitimacy and expertise load. The model fit of a two-factor model should be better than the single factor and second order model. Furthermore, we tested whether the square root of the AVE of each factor was higher than its maximum correlation with other factors, which is an indication of discriminant validity (Fornell and Larcker 1981). The single factor model for all groups combined has an unacceptable fit and, thus, performs less well than our two-factor model: SRMR = .059, RMSEA = .102, TLI = .813, CFI = .866. A second order model has a good fit (SRMR = .031, RMSEA = .053, TLI = .950, CFI = .968). The incremental fit indices for the two-factor model are marginally higher. The square root of the average variance extracted (AVE) are .65 and .66, which is higher than the correlation between the two factors ( $r = .22$ ), indicating that the items share more variance with their own dimension than with other dimensions.

## Organizational types

The expectation of accountability along with expertise and legitimacy of the account holder are theoretically important to measure. Our analysis so far has clarified that the two additional dimensions are distinctive from each other as well as from the original scale for felt (= anticipated) accountability. To further test its relevance, we analysed whether the instrument would also pick up meaningful empirical variance across different groups of respondents. Our first sample of sub-national civil servants consisted of respondents from three types of organizations: local governments, regional governments, and the police. The second sample of national civil servants consisted of respondents from two types of organizations: central government departments and executive agencies. We, thus, tested the differences in felt relational accountability between the five organizational types in our two samples. We compared the expertise and legitimacy of the account holder with the anticipated accountability as measured with the Hochwarter-scale.

Regression analyses with the organizational type as predictor showed that the anticipated accountability had only a limited correlation with organizational types ( $R^2 = .01$ ). Organizational types explained more variance in account holder expertise ( $R^2 = .07$ ) and legitimacy ( $R^2 = .08$ ). Although the anticipation of accountability remains by and large the same across all organizational types, we are now able to observe significant differences in felt relational accountability. This finding emphasizes the convergent validity of the instrument; [Figure 3](#) shows the differences between the organizational types compared to municipalities as a baseline. For anticipated accountabilities, we observe a little less anticipated accountability for police and executive agency personnel. The differences between organizational types for account holder expertise and legitimacy are much larger. We observe that, in comparison to how municipal personnel esteem their political principals, civil servants at the provincial administration gauge the legitimacy and expertise of the provincial government much



**Figure 3.** Comparison of organization types.

Note: estimated mean differences (baseline = municipality) and 95% confidence intervals based on quasi standard errors (Firth 2003).

higher. Police personnel, in contrast, has much lower esteem for the legitimacy and expertise of the ministry of security and justice. This is understandable in the specific context of a tumultuous reorganization of the national police in the Netherlands (Moggré, den Boer, and Fyfe 2017). We also observe that civil servants in ministries feel that their minister is a legitimate account holder, yet they regard their political principal's expertise less high. These findings emphasize the relevance of including expertise and legitimacy when measuring felt relational accountability.

## Discussion and conclusion

In this article, we presented a multidimensional scale to gauge felt relational accountability in public administration. It is a relatively short and flexible scale that can be applied in various contexts. The scale has proved to be valid and reliable across various sectors and levels of government and to be able to measure meaningful variance.

Previous conceptual and empirical analyses of accountability have shown the importance of accountability as a relational concept (Bovens 2007; Brandsma and Schillemans 2012; Bovens, Goodin, and Schillemans 2014). Therefore, we propose to measure felt relational accountability with both the traditional psychological perspective of generic anticipated accountability (Hochwarter et al. 2007), and the two essential dimensions that probe the relational nature of public accountability. Specifically, we add the *perceived expertise* and the *perceived legitimacy* of the account holder as constituting dimensions of felt relational accountability (Lerner and Tetlock 1999; Schillemans 2016). In doing so, the scale developed here provides a more fine-grained insight in an agent's felt relational accountability to some account holder compared with existing measurements, which do not account for the multidimensional and relational nature of the concept (cf. Hochwarter et al. 2007).

The two additional dimensions are relevant for research on the effects of accountability on the behaviour of agents in the public sector. Consider, for instance, how managers in a large public sector organization with a lot of discretionary space behave in the face of a central government department. Whether managers expect to be held accountable for their decisions in the future by the government department or not matters for the manager's decisions and actions. Both psychological as well as public administration research documents that the anticipation of accountability has strong behavioural effects (Lerner and Tetlock 1999; Meijer 2000).

Furthermore, the legitimacy and the expertise of the account holder are crucial to understand responses of actors in organizations. We expect much more constructive, much more *accountable*, responses from the public manager to accountability-demands when the account holder is considered both legitimate, as well as knowledgeable, based on experimental research.

In particular, we expect that the perception of more legitimacy of the account holder will increase the 'willingness to obey' (Simon 1997), as well as the willingness to provide a correct and honest account of behaviour and performance. In lab settings, low perceived legitimacy has been linked to shirking, a decline in intrinsic motivation, and stress (Lerner and Tetlock 1999, 259). At the same time, the expertise of the account holder has been associated with better decisions in lab settings: actors spend more time and cognitive effort on their tasks when they anticipate accountability to an expert (Aleksovska, Schillemans, and Grimmelikhuijsen 2019, Schillemans 2019).

The current instrument facilitates further research among public sector employees in order to validate such experimental findings in public organizations. Doing so is important as accountability is a phenomenon that is *par excellence* subject to socialization, which limits the generalizability of experiments in the behavioural sciences (Juncos and Pomorska 2011). Account holders may even have a strong impact on the behaviour of public managers when they do not expect any accountability yet hold them in high esteem, thus influencing agents through the shadow of hierarchy (cf. Scharpf 1997), or, more specifically in this case, the shadow of authority.

We see a variety of potential applications for this scale. First and most obviously, scholars in accountability could use this scale to gauge the relationship between formal accountability mechanisms and the extent to which these formal mechanisms are felt in public organizations (Overman 2020). Felt relational accountability is, then, a dependent variable. A validated scale can foster knowledge accumulation within the community of accountability scholarship. In addition, our validated scale helps to set up cross-country comparative research. For instance, researchers can use this scale to have a common measure for felt relational accountability across countries and further improve the quality of comparative research on governance in the public sector (Verhoest et al. 2017).

Secondly, felt relational accountability can also be used as an independent variable relating to various salient issues in public administration and to the various promises of accountability (Dubnick and Frederickson 2010). For instance, strong accountability mechanisms are often linked with red tape and administrative overload in our audit society (Power 1999; Halachmi 2014). Now that we can empirically measure felt relational accountability, the relationships and interactions with red tape can be investigated. Do higher levels of felt relational accountability correspond with more red tape? Or is there an inverse relationship? This is relevant as the expansive public administration literature has made it abundantly clear that the mere presence of accountability mechanisms does not necessarily, to put it mildly, lead to desirable outcomes, given the prevalence of many accountability failures (Halachmi 2014; Koppell 2005).

Thirdly, we see potential applications of this scale beyond the discipline of public administration research. Our point of departure was an existing unidimensional scale of accountability often used in the behavioural sciences (Hochwarter et al. 2007; Hall, Frink, and Ronald Buckley 2017). We argue that disciplines such as (social) psychology can refine their experimental tests by taking into account the relational nature of accountability in public administration. Of course, not all items might be applicable in a more generic context, but our scale does show that a richer conceptualization and operationalization of accountability is needed. This could then be used to assess behavioural consequences of specific dimensions of accountability on micro-level behaviours. Hence the validated scale can be a way to increase cross-fertilization between disciplines and in this way strengthen what has been labelled behavioural public administration (Grimmelikhuijsen et al. 2017).

## Notes

1. A Web of Science query for the topic accountability in SSCI listed public administration journals yielded the following numbers of articles per year: 2011 – 135; 2012 – 171; 2013 – 158; 2014 – 187; 2015 – 165; 2016 – 189; 2017 – 166.
2. Items include: I am held very accountable for my actions at work; I often have to explain why I do certain things at work; Top management holds me accountable for all of my decisions; If



things at work do not go the way that they should; I will hear about it from top management; To a great extent; the success of my immediate work group rests on my shoulders; The jobs of many people at work depend on my success or failures; In the grand scheme of things; my efforts at work are very important; co-workers; subordinates; and bosses closely scrutinize my efforts at work. The internal consistency in the current sample ( $\alpha=.73$ ) is identical to prior tests (Hall and Ferris 2011).

3. The term felt accountability is commonly accepted in the behavioural sciences and, therefore, the authors chose to use the existing term for consistency, rather than alternatives including experienced or perceived accountability.

## Disclosure statement

No potential conflict of interest was reported by the author(s).

## Funding

This work was supported by the Nederlandse Organisatie voor Wetenschappelijk Onderzoek [425-14-008 'Calibrating Public Accountability'].

## Notes on contributors

*Sjors Overman* is assistant professor at Utrecht University, School of Governance. He studies governance of public organizations, felt accountability in the public sector, and effects of public sector reforms with a focus on quantitative analysis.

*Thomas Schillemans* is Professor in Accountability, Behaviour and Governance at Utrecht University, School of Governance. His research focuses on the interactions of public sector organizations with various relevant stakeholders from their environment. He specializes in public accountability, public sector governance, trust & control and the role of the media in policy implementation.

*Stephan Grimmelikhuijsen* is associate professor at Utrecht University, School of Governance. Core themes in his research are: transparency, citizen trust in government, technology in public administration, experimental methodology, behavioural public administration.

## ORCID

*Sjors Overman*  <http://orcid.org/0000-0002-0620-0682>

*Stephan Grimmelikhuijsen*  <http://orcid.org/0000-0002-1553-6065>

## References

- Aleksovska, M., T. Schillemans, and S. Grimmelikhuijsen. 2019. "Lessons from Five Decades of Experimental and Behavioral Research on Accountability: A Systematic Literature Review." *Journal of Behavioral Public Administration* 2 (2): 2. doi:10.30636/jbpa.22.66.
- Alford, J., S. Douglas, K. Geuijen, and P. T. Hart. 2017. *Ventures in Public Value Management: Introduction to the Symposium*. *Public Management Review* 19 (5): 589–604.
- Bach, T., S. van Thiel, G. Hammerschmid, and R. Steiner. 2016. "Administrative Tradition and Management Reforms: A Comparison of Agency Chief Executive Accountability in Four Continental Rechtsstaat Countries." *Public Management Review* 19 (6): 765–784. doi:10.1080/14719037.2016.1210205.
- Bovens, M. 19 2005. "Public Accountability." In *The Oxford Handbook of Public Management*, edited by E. Ferlie, L. E. Lynn, and C. Pollitt, 182–208. Oxford; New York: Oxford University Press.
- Bovens, M. 2007. "Analysing and Assessing Accountability: A Conceptual Framework." *European Law Journal* 13 (4): 447–468. doi:10.1111/j.1468-0386.2007.00378.x.

- Bovens, M. 2010. "Two Concepts of Accountability: Accountability as a Virtue and as a Mechanism." *West European Politics* 33 (5): 946–967. doi:10.1080/01402382.2010.486119.
- Bovens, M., T. Schillemans, and R. E. Goodin. 19 2014. "Public Accountability." In *The Oxford Handbook of Public Accountability*, edited by M. Bovens, R. E. Goodin, and T. Schillemans, 1–20. Oxford: Oxford University Press.
- Bovens, M., and A. Wille. 2016. *The EU Accountability Index: Assessing the Accountability Powers of Two EU Watchdogs*. Utrecht: EGPA Annual Conference.
- Bovens, M., R. E. Goodin, and T. Schillemans. 2014. *The Oxford Handbook of Public Accountability*. Oxford: Oxford University Press.
- Bovens, M., T. Schillemans, and P. T. Hart. 2008. "Does Public Accountability Work? An Assessment Tool." *Public Administration* 86 (1): 225–242. doi:10.1111/j.1467-9299.2008.00716.x.
- Brandsma, G. J. 19 2014. "Quantitative Analysis." In *The Oxford Handbook of Public Accountability*, edited by M. Bovens, R. E. Goodin, and T. Schillemans, 143–158. Oxford: Oxford University Press.
- Brandsma, G. J., and T. Schillemans. September 2012. "The Accountability Cube: Measuring Accountability." *Journal of Public Administration Research and Theory*: mus034. doi:10.1093/jopart/mus034.
- Brehm, J. O., and S. Gates. 1999. *Working, Shirking, and Sabotage: Bureaucratic Response to a Democratic Public*. Ann Arbor: University of Michigan Press.
- Carman, J. G. 2008. "Nonprofits, Funders, and Evaluation: Accountability in Action." *The American Review of Public Administration*. <http://arp.sagepub.com/content/early/2008/06/13/0275074008320190.short>.
- Carpenter, D., and G. A. Krause. 2015. "Transactional Authority and Bureaucratic Politics." *Journal of Public Administration Research and Theory* 25 (1): 5–25. doi:10.1093/jopart/muu012.
- Coleman, J. S. 1990. *Foundations of Social Theory*. Cambridge, Mass: Harvard University Press.
- Curtin, D. 2007. "Holding (Quasi-)autonomous EU Administrative Actors to Public Account." *European Law Journal* 13 (4): 523–541. doi:10.1111/j.1468-0386.2007.00382.x.
- DeVellis, R. F. 2009. *Scale Development: Theory and Applications*. 2nd ed. Thousand Oaks, CA: Sage Publ.
- Donahue, J. D., and M. H. Moore. 2012. *Ports in a Storm: Public Management in a Turbulent World*. Washington, DC: Brookings Institution Press.
- Dubnick, M., and B. Romzek. 1993. "Accountability and the Centrality of Expectations in American Public Administration." In *Research in Public Administration*, edited by J. Perry, 37–78. Greenwich: JAI, Press.
- Dubnick, M. J. 19 2014. "Accountability as a Cultural Keyword." In *The Oxford Handbook of Public Accountability*, edited by M. Bovens, R. E. Goodin, and T. Schillemans, 23–38. Oxford: Oxford University Press.
- Dubnick, M. J., and H. George Frederickson. 2010. "Accountable Agents: Federal Performance Measurement and Third-Party Government." *Journal of Public Administration Research and Theory* 20 (suppl\_1): i143–59. doi:10.1093/jopart/mup039.
- Dubnick, M. J., and H. George Frederickson. 2011. *Accountable Governance: Problems and Promises*. Armonk, NY: ME Sharpe.
- Ebrahim, A. 2003. "Making Sense of Accountability: Conceptual Perspectives for Northern and Southern Nonprofits." *Nonprofit Management and Leadership* 14 (2): 191–212. doi:10.1002/nml.29.
- Firth, D. 2003. "Overcoming the Reference Category Problem in the Presentation of Statistical Models." *Sociological Methodology* 33 (1): 1–18. doi:10.1111/j.0081-1750.2003.t01-1-00125.x.
- Ford, J. B., and M. Zelditch Jr. 1988. "A Test of the Law of Anticipated Reactions." *Social Psychology Quarterly* 164–171. doi:10.2307/2786838.
- Fornell, C., and D. F. Larcker. 1981. "Evaluating Structural Equation Models with Unobservable Variables and Measurement Error." *Journal of Marketing Research* 18 (1): 39–50. doi:10.2307/3151312.
- Friedrich, C. J. 1963. *Man and His Government. An Empirical Theory of Politics*. New York: McGraw Hill.
- Frink, D. D., and R. J. Klimoski. 1998. "Toward a Theory of Accountability in Organizations and Human Resource Management." In *Research in Personnel and Human Resources Management*, edited by G. R. Ferris, 1–51. Vol. 16. US: Elsevier Science/JAI Press.
- Gailmard, S. 2014. "Accountability and Principal-Agent Theory." In *The Oxford Handbook of Public Accountability*, edited by M. Bovens, R. E. Goodin, and T. Schillemans, 90–105. Oxford: Oxford University Press.

- Grimmelikhuijsen, S., S. Jilke, A. L. Olsen, and L. Tummers. 2017. "Behavioral Public Administration: Combining Insights from Public Administration and Psychology." *Public Administration Review* 77 (1): 45–56. doi:10.1111/puar.12609.
- Halachmi, A. 19 2014. "Accountability Overloads." In *The Oxford Handbook of Public Accountability*, edited by M. Bovens, R. E. Goodin, and T. Schillemans, 560–573. Oxford: Oxford University Press.
- Hall, A. T., D. D. Frink, and M. Ronald Buckley. 2017. "An Accountability Account: A Review and Synthesis of the Theoretical and Empirical Research on Felt Accountability: Accountability." *Journal of Organizational Behavior* 38 (2): 204–224. doi:10.1002/job.2052.
- Hall, A. T., and G. R. Ferris. 2011. "Accountability and Extra-Role Behavior." *Employee Responsibilities and Rights Journal* 23 (2): 131–144. doi:10.1007/s10672-010-9148-9.
- Harari, M. B., and C. W. Rudolph. 2017. "The Effect of Rater Accountability on Performance Ratings: A Meta-Analytic Review." *Human Resource Management Review* 27 (1): 121–133. Using Meta-analysis to Enhance our Understanding of Human Resource Management. doi:10.1016/j.hrmr.2016.09.007.
- Hinkin, T. R. 1998. "A Brief Tutorial on the Development of Measures for Use in Survey Questionnaires." *Organizational Research Methods* 1 (1): 104–121. doi:10.1177/109442819800100106.
- Hochwarter, W. A., G. R. Ferris, M. B. Gavin, P. L. Perrewé, A. T. Hall, and D. D. Frink. 2007. "Political Skill as Neutralizer of Felt Accountability—job Tension Effects on Job Performance Ratings: A Longitudinal Investigation." *Organizational Behavior and Human Decision Processes* 102 (2): 226–239. doi:10.1016/j.obhdp.2006.09.003.
- Hochwarter, W. A., P. L. Perrewé, A. T. Hall, and G. R. Ferris. 2005. "Negative Affectivity as a Moderator of the Form and Magnitude of the Relationship between Felt Accountability and Job Tension." *Journal of Organizational Behavior* 26 (5): 517–534. doi:10.1002/job.324.
- Hong, S., S. H. Kim, and J. Son. 2019. "Bounded Rationality, Blame Avoidance, and Political Accountability: How Performance Information Influences Management Quality." *Public Management Review*. Online first. doi:10.1080/14719037.2019.1630138.
- Hooghe, L., and M. Gary. 2003. "Unraveling the Central State, but How? Types of Multi-level Governance." *American Political Science Review* 97 (2): 233–243.
- Hooper, D., J. Coughlan, and M. Mullen. 2008. "Structural Equation Modelling: Guidelines for Determining Model Fit." *Electronic Journal of Business Research Methods* 6 (1): 53–60.
- Hu, L.-T., and P. M. Bentler. 1998. "Fit Indices in Covariance Structure Modeling: Sensitivity to Underparameterized Model Misspecification." *Psychological Methods* 3 (4): 424. doi:10.1037/1082-989X.3.4.424.
- Juncos, A. E., and K. Pomorska. 2011. "Invisible and Unaccountable? National Representatives and Council Officials in EU Foreign Policy." *Journal of European Public Policy* 18 (8): 1096–1114. doi:10.1080/13501763.2011.615197.
- Kim, S., W. Vandenberg, B. E. Wright, L. B. Andersen, F. P. Cerase, R. K. Christensen, C. Desmarais, et al. 2013. "Investigating the Structure and Meaning of Public Service Motivation across Populations: Developing an International Instrument and Addressing Issues of Measurement Invariance." *Journal of Public Administration Research and Theory* 23 (1): 79. doi:10.1093/jopart/mus027.
- Kline, R. B. 2015. *Principles and Practice of Structural Equation Modeling*. 4th ed. New York: Guilford Publications.
- Klingner, D. E., J. Nalbandian, and B. S. Romzek. 2002. "Politics, Administration, and Markets: Conflicting Expectations and Accountability." *The American Review of Public Administration* 32 (2): 117–144. doi:10.1177/02774002032002001.
- Koop, C. 2011. "Explaining the Accountability of Independent Agencies: The Importance of Political Salience." *Journal of Public Policy* 31 (2): 209–234. doi:10.1017/S0143814X11000080.
- Koop, C. 2014. "Theorizing and Explaining Voluntary Accountability." *Public Administration* 92 (3): 565–581. doi:10.1111/padm.12058.
- Koppell, J. G. S. 2005. "Pathologies of Accountability: ICANN and the Challenge of 'Multiple Accountabilities Disorder'." *Public Administration Review* 65 (1): 94–108. doi:10.1111/j.1540-6210.2005.00434.x.
- Kumar, S., N. Pandey, and A. Haldar. 2020. "Twenty Years of Public Management Review (PMR): A Bibliometric Overview." *Public Management Review* 1–21. doi:10.1080/14719037.2020.1721122.
- Lerner, J. S., and P. E. Tetlock. 1999. "Accounting for the Effects of Accountability." *Psychological Bulletin* 125 (2): 255–275. doi:10.1037/0033-2909.125.2.255.

- Lowe, I. 19 2007. "The Research Community." In *Silencing Dissent. How the Australian Government Is Controlling Public Opinion and Stifling Debate*, edited by C. Hamilton and S. Maddison, 60–77. Crow's Nest, Australia: Allen & Unwin.
- Lupson, J., and D. Partington. 2011. "Individual Civil Servants' Conceptions Of Accountability: A Preliminary Study." *Public Management Review* 13 (7): 895–918. doi:10.1080/14719037.2011.589609.
- MacCallum, R. C., M. Roznowski, and L. B. Necowitz. 1992. "Model Modifications in Covariance Structure Analysis: The Problem of Capitalization on Chance." *Psychological Bulletin* 111 (3): 490. doi:10.1037/0033-2909.111.3.490.
- Meijer, A. 2000. "Anticipating Accountability Processes." *Archives and Manuscripts* 28 (1): 52.
- Mero, N. P., R. M. Guidice, and S. Werner. 2014. "A Field Study of the Antecedents and Performance Consequences of Perceived Accountability." *Journal of Management* 40 (6): 1627–1652. doi:10.1177/0149206312441208.
- Miller, J. P. 1970. "Social-psychological Implications of Weber's Model of Bureaucracy: Relations among Expertise, Control, Authority, and Legitimacy." *Social Forces* 49 (1): 91–102. doi:10.2307/2575742.
- Moggré, M., M. den Boer, and N. R. Fyfe. 2017. "Police Leadership in Times of Transition." *Policing: A Journal of Policy and Practice* 12 (4): 388–397.
- Mulgan, R. 2000. "'Accountability': An Ever-Expanding Concept?." *Public Administration* 78 (3): 555–573. doi:10.1111/1467-9299.00218.
- Mulgan, R. G. 2003. *Holding Power to Account: Accountability in Modern Democracies*. Basingstoke: Palgrave Macmillan.
- Olsen, J. P. 2013. "The Institutional Basis of Democratic Accountability." *West European Politics* 36 (3): 447–473. doi:10.1080/01402382.2012.753704.
- Olsen, J. P. 2017. "Democratic Accountability and the Terms of Political order." *European Political Science Review* 9 (4): 519–537. doi:10.1017/S1755773916000084.
- Overman, S. 2020. "Aligning Accountability Arrangements for Ambiguous Goals. The Case of Museums." *Public Management Review*: 1–21. Online first. doi:10.1080/14719037.2020.1722210.
- Overman, S., M. Busuioic, and M. Wood. 2020. "A Multidimensional Reputation Barometer for Public Agencies: A Validated Instrument." *Public Administration Review*. Online first. doi:10.1111/puar.13158.
- Overman, S., M. Van Genugten, and S. Van Thiel. 2015. "Accountability After Structural Disaggregation: Comparing Agency Accountability Arrangements." *Public Administration* 93 (4): 1102–1120. doi:10.1111/padm.12185.
- Page, S. 2006. "The Web of Managerial Accountability the Impact of Reinventing Government." *Administration & Society* 38 (2): 166–197. doi:10.1177/0095399705285990.
- Pandey, S. K., and P. G. Scott. 2002. "Red Tape: A Review and Assessment of Concepts and Measures." *Journal of Public Administration Research and Theory* 12 (4): 553–580. doi:10.1093/oxfordjournals.jpart.a003547.
- Patil, S. V., F. Vieider, and P. E. Tetlock. 19 2014. "Process versus Outcome Accountability." In *The Oxford Handbook of Public Accountability*, edited by M. Bovens, R. E. Goodin, and T. Schillemans, 69–89. Oxford: Oxford University Press.
- Peng, A. C., J. Dunn, and D. E. Conlon. 2015. "When Vigilance Prevails: The Effect of Regulatory Focus and Accountability on Integrative Negotiation Outcomes." *Organizational Behavior and Human Decision Processes* 126: 77–87. doi:10.1016/j.obhdp.2014.10.008.
- Power, M. 1999. *The Audit Society: Rituals of Verification*. Oxford; New York: Oxford University Press.
- Presthus, R. V. 1960. "Authority in Organizations." *Public Administration Review* 20 (2): 86–91. doi:10.2307/973885.
- Raykov, T. 2001. "Estimation of Congeneric Scale Reliability Using Covariance Structure Analysis with Nonlinear Constraints." *British Journal of Mathematical and Statistical Psychology* 54 (2): 315–323. doi:10.1348/000711001159582.
- Romzek, B. S., K. LeRoux, J. Johnston, R. J. Kempf, and J. S. Piatak. 2014. "Informal Accountability in Multisector Service Delivery Collaborations." *Journal of Public Administration Research and Theory* 24 (4): 813–842. doi:10.1093/jopart/mut027.
- Romzek, B. S., K. LeRoux, and J. M. Blackmar. 2012. "A Preliminary Theory of Informal Accountability among Network Organizational Actors." *Public Administration Review* 72 (3): 442–453. doi:10.1111/j.1540-6210.2011.02547.x.

- Romzek, B. S., and M. J. Dubnick. 1987. "Accountability in the Public Sector: Lessons from the Challenger Tragedy." *Public Administration Review* 47 (3): 227–238. doi:10.2307/975901.
- Rossee, Y. 2012. "Lavaan: An R Package for Structural Equation Modeling." *Journal of Statistical Software* 48 (2): 2. doi:10.18637/jss.v048.i02.
- Scharpf, F. W. 1997. *Games Real Actors Play. Actor-Centered Institutionalism in Policy Research*. Boulder, Colo: Westview Press.
- Schillemans, T. 2011. "Does Horizontal Accountability Work? Evaluating Potential Remedies for the Accountability Deficit of Agencies." *Administration & Society* 43 (4): 387–416. doi:10.1177/0095399711412931.
- Schillemans, T. 2016. "Calibrating Public Sector Accountability: Translating Experimental Findings to Public Sector Accountability." *Public Management Review* 18 (9): 1400–1420. doi:10.1080/14719037.2015.1112423.
- Schillemans, T., and M. Busuioic. 2015. "Predicting Public Sector Accountability: From Agency Drift to Forum Drift." *Journal of Public Administration Research and Theory* 25 (1): 191–215.
- Scott, W. R. 1965. "Reactions to Supervision in a Heteronomous Professional Organization." *Administrative Science Quarterly* 65–81. doi:10.2307/2391650.
- Simon, H. A. 1997. *Administrative Behavior: A Study of Decision-Making Processes in Administrative Organizations*. Vol. 4. ed. New York, NY: Free Press.
- Strom, K. 2000. "Delegation and Accountability in Parliamentary Democracies." *European Journal of Political Research* 37 (3): 261–290. doi:10.1023/A:1007064803327.
- Suchman, M. C. 1995. "Managing Legitimacy: Strategic and Institutional Approaches." *Academy of Management Review* 20 (3): 571–610. doi:10.5465/amr.1995.9508080331.
- Tetlock, P. E. 1992. "The Impact of Accountability on Judgment and Choice: Toward a Social Contingency Model." In *Advances in Experimental Social Psychology*, Vol. 25, 331–376. Elsevier.
- Tetlock, P. E., L. Skitka, and R. Boettger. 1989. "Social and Cognitive Strategies for Coping with Accountability: Conformity, Complexity, and Bolstering." *Journal of Personality and Social Psychology* 57 (4): 632. doi:10.1037/0022-3514.57.4.632.
- Tyler, T. R. 1997. "The Psychology of Legitimacy: A Relational Perspective on Voluntary Deference to Authorities." *Personality and Social Psychology Review* 1 (4): 323–345. doi:10.1207/s15327957pspr0104\_4.
- van Loon, N. M., P. L. M. Leisink, E. Knies, and G. A. Brewer. 2016. "Red Tape: Developing and Validating a New Job-Centered Measure." *Public Administration Review* 76 (4): 662–673. doi:10.1111/puar.12569.
- Verhoest, K., J. Wynen, W. Vandenabeele, and S. Van De Walle. 19 2017. "Challenges for Large-Scale International Comparative Survey-Based Research in Public Administration." In *The Palgrave Handbook of Public Administration and Management in Europe*, edited by E. Ongaro and S. Van Thiel, 1147–1168. London: Palgrave Macmillan UK. doi:10.1057/978-1-137-55269-3.
- Verhoest, K., P. G. Roness, B. Verschuere, K. Rubecksen, and M. Muiris. 2010. *Autonomy and Control of State Agencies*. Houndmills, Basingstoke, Hampshire: Palgrave Macmillan.
- Waldo, D. 1948. *The Administrative State*. New York: Ronald Press Company.
- Wong, W., and E. Welch. 2004. "Does E-Government Promote Accountability? A Comparative Analysis of Website Openness and Government Accountability." *Governance* 17 (2): 275–297. doi:10.1111/j.1468-0491.2004.00246.x.
- Yang, K. 2012. "Further Understanding Accountability in Public Organizations: Actionable Knowledge and the Structure–agency Duality." *Administration & Society* 44 (3): 255–284. doi:10.1177/0095399711417699.
- Yang, K. 19 2014. "Qualitative Analysis." In *The Oxford Handbook of Public Accountability*, edited by M. Bovens, R. E. Goodin, and T. Schillemans, 159–176. Oxford: Oxford University Press.