Nina M. van Loon
Aarhus University, Denmark
Peter L. M. Leisink
Eva Knies
Utrecht University, The Netherlands

Gene A. Brewer
The University of Georgia

Red Tape:
Developing and Validating a New Job-Centered Measure

Nina M. van Loon is assistant professor in the Department of Political Science and Government, Aarhus University, Denmark. Her research focuses on the performance of public services and how this can be improved, including research on red tape, public service motivation, and performance management.
E-mail: nina@ps.au.dk

Peter L. M. Leisink is professor of public administration and organization science in the Utrecht University School of Governance, The Netherlands. His research interests are in the areas of management and organization of service organizations, the contribution of strategic human resource management to public value creation in public and nonprofit organizations, leadership and motivation in public organizations, and public service performance under austerity. His research has been published in various international journals.
E-mail: p.l.m.leisink@uu.nl

Eva Knies is associate professor in organizational science and human resource management in the Utrecht University School of Governance, The Netherlands. Her research interests include public value creation, public service performance, human resource management in the public sector (in particular education and health care), and public leadership. Her research has been published in various international journals in the fields of public administration, public management, and human resource management.
E-mail: e.knies@uu.nl

Gene A. Brewer is a longtime faculty member in The University of Georgia School of Public and International Affairs, Department of Public Administration and Policy. He is also visiting professor of public management in the Utrecht University School of Governance in The Netherlands and guest fellow of public management in the Public Governance Institute at Catholic University in Leuven, Belgium.
E-mail: geneabrewer@uga.edu

Abstract: *Red tape studies typically focus on burdensome rules that have negative effects on organizations, as perceived by managers. The one-item general red tape scale is representative of this approach. However, scholars have called for improved measures that address the scale's shortcomings. This article introduces a new measurement scale that features (1) red tape as a two-dimensional construct that includes compliance burden and lack of functionality and (2) a job-centered approach that measures red tape as experienced by employees in their jobs rather than more generally in the organization. A set of survey questions derived from interviews with government employees was validated using data from 1,203 government employees. The findings indicate that the two-dimensional job-centered red tape scale is reliable and valid. The authors conclude that this measure can improve research and be used by managers for a "quick scan" to detect the location and severity of red tape.*

Practitioner Points
- Public organizations are under pressure to improve public services. Understanding red tape in the individual work situations of employees helps in creating better circumstances for employees to perform by providing insight into how the rules affect individual employees' work.
- The job-centered red tape scale can be applied by organizations that want a "quick scan" to gain insight into how many employees perceive the rules they have to work with as lacking functionality and/or as imposing a high compliance burden.
- The two-dimensional red tape scale focuses on rules as perceived by employees in their own job, making it possible to compare organizational units (departments, teams) *within* the organization and, consequently, to improve the working conditions.

The concept of red tape is relevant to both scholars and practitioners because of its presumed prevalence in government and its destructive impact on performance. Bozeman defined organizational red tape as "rules, regulations, and procedures that remain in force and entail a compliance burden for the organization but have no efficacy for the rules' functional object" (1993, 283). Since his influential publication, research on red tape has produced valuable insights, but several authors (including Bozeman himself) have raised questions about current measures of the concept (Bozeman and Feeney 2011; Brewer and Walker 2012; Coursey and Pandey 2007; Feeney 2012; Pandey and Scott 2002).

Several studies have used a one-item measure called the general red tape (GRT) scale that asks managers to indicate the overall level of red tape in their organization using a 10-point Likert-type scale (Bozeman and Feeney 2011; Rainey, Pandey, and Bozeman 1995). Although this simple measure has

appeal, researchers question whether it measures the concept of red tape well (Feeney 2012). Other studies have tried to zoom in and measure red tape in management subsystems such as personnel, budgeting and procurement, and information technology (DeHart-Davis and Pandey 2005; Pandey and Kingsley 2000; Pandey, Coursey, and Moynihan 2007; Scott and Pandey 2005; Walker and Brewer 2008). Some researchers have sought to gauge red tape by assessing the difficulties managers face and the time they spend complying with rules (Chen and Williams 2007; Pandey and Bretschneider 1997; Pandey, Coursey, and Moynihan 2007; Scott and Pandey 2005). Others have proposed measuring red tape among stakeholder groups based on the assumption that such groups are likely to perceive rules differently (Bozeman and Feeney 2011; Brewer and Walker 2010a). All of these approaches have merits but also shortcomings. Reflecting on the current state of red tape measurement, Bozeman and Feeney (2011) urge researchers to develop additional measures to advance research on the topic.

This article responds to this call. We identify three specific measurement issues that have impeded research progress, and we address these issues by developing a new job-centered measure of red tape that focuses on rules that affect employees in the performance of their duties. This multiple-item scale incorporates two dimensions of red tape that are prominent in the literature but have not been included together in contemporary measurement schemes: a rule's compliance burden and lack of functionality.

The article makes three significant contributions to the literature on red tape. First, current measures of red tape typically focus on the compliance burden of rules and pay less explicit attention to the efficacy of the rules in achieving their functional objective, although the latter is a crucial element in Bozeman's (1993) definition. As a consequence, these studies typically overlook that burdensome rules are not red tape per se (Feeney 2012). In government organizations, administrative rules may have purposes other than efficacy (which is a combination of efficiency and effectiveness), such as guaranteeing accountability, dispersing benefits equitably, or protecting the public interest (Bozeman 1993; Frederickson 2005; Kaufman 1977). Thus, a rule that is burdensome may have the functional objective of safeguarding some stakeholders' interests—a point well made by those who study stakeholder red tape (Bozeman and Feeney 2011; Brewer and Walker 2010a). As Kaufman observed, "one person's red tape is another's sacred protection" (1977, 61). Our measure of red tape includes both the compliance burden *and* lack of rule functionality specified in Bozeman's (1993) definition. Both elements have deep roots in the literature on red tape.

Second, many studies ask respondents to report on the level of red tape in their organizations (Bozeman and Feeney 2011). Yet it is very difficult, if not impossible, for respondents to take stock of the organization's rule environment and summarize its impact on organizational effectiveness. In other words, respondents may not be familiar with many organizational rules or their impact on organizational performance, which is a very distant outcome for most employees. Moreover, where the respondent sits in the organization likely determines what he or she sees. For example, budget analysts may think of procurement and expenditure rules when assessing the organization's rule environment, and they may conflate its financial performance with its overall effectiveness. We contend that employees can provide a more valid assessment of the rules they encounter in their own jobs, so our measure is based on their perceptions of the rules affecting their *own* work situation. Thus, we develop a measure of job-centered red tape that supplements existing measures.

Third, consonant with the interest in organizational red tape, previous studies have asked managers to rate the extent of red tape generally or in management subsystems such as personnel, budget and procurement, information systems, and communication. Red tape affecting the domain of policy development and implementation has not been examined. Moreover, red tape in management subsystems such as personnel and finance is often measured through items that feature the type of tasks that *managers* may perform, such as firing a poor performer or rewarding a

good one, but do not include the vastly larger number of rules that managers and nonmanagers alike frequently encounter in their work, such as requesting permission to attend a training session (personnel) or submitting a statement of expenses for reimbursement (finance). When measuring job-centered red tape, therefore, it is necessary to develop a measure that is applicable to all job domains and includes tasks that all employees may perform. This is what we have done in the present study.

This article contributes to the study of red tape by developing a measure of job-centered red tape that supplements existing measures of organizational and stakeholder red tape and addresses their shortcomings as described earlier. The development of a job-centered red tape scale involved two steps. First, interviews were conducted to explore how employees experience rules in their jobs, and this information was used to develop a battery of measurement items. Second, these items were validated using survey data from 1,203 employees in the Dutch central government.

The article proceeds with an overview of the literature on red tape measurement focusing on the measurement issues mentioned earlier, followed by a description of our methods and results on item development, scale construction, and validation. Finally, the value of the job-centered red tape scale for theory and practice is discussed.

Theoretical Framework

Unlike good rules, organizational red tape is seen as an inherently negative concept that frustrates employees in reaching their goals (Brewer and Walker 2010a; Moynihan and Pandey 2007; Pandey and Scott 2002) and cannot lead to anything positive (Bozeman 1993; Scott and Pandey 2000). Conceptually, it may make sense to define rules as red tape by the negative consequences they have, but for measuring red tape, this creates problems. Rules, which are the cause, and negative effects, which are the consequences for the organization, are distinct phenomena. The cause must precede the effect; otherwise, one could only distinguish red tape retrospectively—after it has demonstrated negative effects. This is a paralyzing limitation for red tape researchers. Public management is a design science that aims to understand causal mechanisms (such as the impact of red tape on performance) and use this knowledge to improve outcomes. If researchers are unable to recognize the cause before the effects set in, they cannot make causal attributions or offer design-based improvements from such evidence. This study takes another approach by using Bozeman's (1993) original definition of red tape and operationalizing it at the individual employee's job level, focusing on routine job activities. We define red tape as rules that employees perceive as burdensome and not helpful in achieving the rules' functional objective in their respective job. We call this *job-centered red tape*.

Although Bozeman (1993) used both compliance burden and lack of efficacy for the rules' functional objective to define red tape, the latter characteristic has received little explicit attention in subsequent research. Perhaps this is because rules do not always include an explicit statement of their objective (Bozeman 2012, 254) or because the objectives can change over time (Brewer and Walker 2012). This neglect of the rules' functional objective is

> Rules, which are the cause, and negative effects, which are the consequences for the organization, are distinct phenomena.

illustrated by studies that have examined red tape in management systems through questionnaire items such as "the rules make it hard to fire a poorly performing manager" or "…reward a good manager" (e.g., Brewer and Walker 2010a, 2010b; Brewer et al. 2012; Coursey and Pandey 2007; Moynihan, Wright, and Pandey 2012; Pandey, Coursey, and Moynihan 2007; Pandey and Scott 2002; Scott and Pandey 2005). The explicit use of the words "make it hard" refers to the compliance burden while introducing a possible lack of efficacy without making the rules and their functional objectives explicit. This lack of precision results in an incomplete measure (that is, it does not rigorously incorporate both elements of the conceptual definition), and it provides insufficient structure for respondents, thus increasing the probability of measurement error. For example, Feeney (2012) showed that the GRT scale adequately captures rules that impede organizational effectiveness, but it does not compel respondents to consider other legitimate purposes of rules such as fairness, equity, and accountability.

Our measure of red tape improves on previous measures by incorporating both characteristics of the conceptual definition. First, rules perceived as red tape entail a compliance burden: the employee feels burdened in executing the rule because it requires excessive time or energy, or it is complex or frustrating. It should be noted that this is not a sufficient condition to declare a rule "red tape" because some burdensome rules are actually worth the trouble of executing. Hospital safety guidelines are good examples. A second characteristic is necessary to distinguish red tape from burdensome rules in general. This characteristic is described by Bozeman as having "no efficacy for the rules' functional object" but should be understood as the perceived *degree* of lack of effectiveness rather than as "absolutely no functionality" (Bozeman 2012, 257). In other words, rules can be fully functional, partially functional, or lacking functionality. A rule created to increase accountability that requires much paperwork is not necessarily red tape because it may be worth the effort in terms of increasing accountability. Only when the rule fails to fully serve the purpose of accountability does it become red tape—and then often as a matter of degree. Such a rule might eventually require a few redundant or obsolete reports and begin its transformation into a mild form of red tape.

> Only when the rule fails to fully serve the purpose of accountability does it become red tape—and then often as a matter of degree.

Figure 1 maps the two dimensions of rules found in the conceptual definition, which together form the basis for determining whether rules are red tape: the perceived degree of inefficacy for the rule's functional objective, hereafter referred to as lack of functionality, and the degree of compliance burden. Lack of functionality refers to a rule that is not achieving its functional objective—that is, not serving its intended purpose. Compliance burden can be described as the excessive or unnecessary amount of time, energy, or other resources spent in executing a rule.

A rule that is perceived to entail a high compliance burden and lacks functionality qualifies as *red tape*. Rules that have a high compliance burden but achieve the rules' functional objective can be described as *necessary bureaucracy*. Rules that do not entail a high compliance burden but lack functionality can be labeled *unnecessary rules*. These rules are in force but do not serve their intended purpose. They are thus excessive—even though they do not impose a compliance burden on employees. Finally, rules that have a low compliance burden and achieve their functional objective can be seen as *high-quality rules*. Both high-quality rules and necessary bureaucracy score high on functionality, and in this respect, they could be seen as related to green tape, defined by DeHart-Davis (2009, 362) as effective organizational rules. However, green tape focuses on organizational level rules, does not refer to compliance burden, and is delineated by other attributes, making it a distinct area of research.

Job-Centered Red Tape
Most red tape studies ask managers to rate the level or volume of red tape in their organization or in a variety of management subsystems. One may question the validity of such assessments because respondents are being asked to report on the burden and functionality of rules they may not be familiar with. Such measures do not capture how red tape affects public employees personally. The validity of red tape assessments can be increased by narrowing the scope of inquiry to only those rules that employees have direct knowledge of and experience in their jobs. Studying how rule burden and functionality are perceived by employees in their jobs rather than asking them to speculate on the far greater number of organization-wide rules should improve measurement validity (DeVellis 2003).

In addition, many rules are implemented by interorganizational networks, and employees may experience difficulties in their work because of those rules (Brewer 2013). The validity of red tape assessments can thus be increased by expanding the scope of inquiry to include rules that affect the employee's work but originate in these larger network settings rather than in the employee's home organization. This, too, should help improve the validity of red tape assessments.

Focusing on the individual's job domain allows us to examine red tape in various functions or services and at different levels of the hierarchy. For example, researchers could gauge red tape in various departments; they could see how red tape affects policy analysis and policy-making activities, which has not been studied in prior research; and they could develop keener insights on how red tape affects the work of street-level bureaucrats. In developing our measure, we thus focus on red tape perceived by the employee in his or her own job domain. This is a specific type of stakeholder red tape perceived by employees within the organization (Bozeman and Feeney 2011; Brewer and Walker 2010b). Prior studies of stakeholder red tape have typically considered employees to be a single stakeholder group with a shared perspective (e.g., Feeney and Bozeman 2009; Quratulain and Khan 2015). This study provides a more detailed and nuanced view of red tape by taking a job-centered approach and recognizing that employees may have different perceptions.

Red Tape Pertaining to Various Job Activities
Most prior studies have focused on managers and their perceptions of red tape (Brewer and Walker 2010b; Coursey and Pandey 2007; Moynihan and Pandey 2007; Pandey and Scott 2002; Pandey,
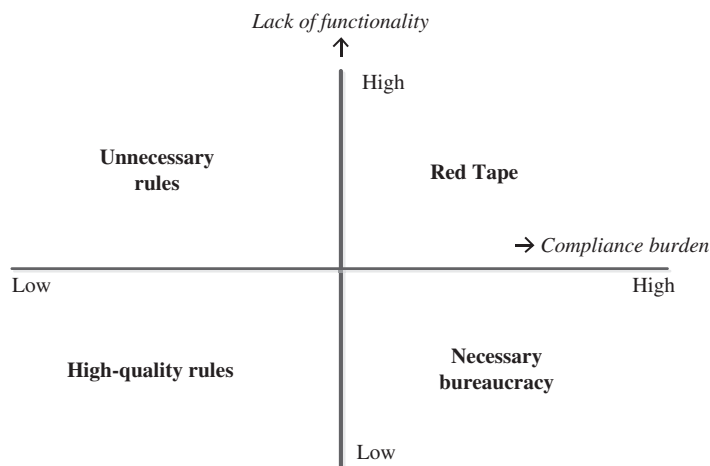
Figure 1 Two Dimensions of Rules: Perceived Lack of Functionality and Compliance Burden

Coursey, and Moynihan 2007). We agree with DeHart-Davis (2009) and Walker and Brewer (2008) that the focus on managers is too narrow and that employees without managerial responsibility can also provide valuable information. In fact, frontline employees probably know more about their specific jobs and the rules that impact their work than do managers. It is also highly likely that nonmanagerial employees encounter more red tape than managers. Moving down the hierarchy, each level of management adds additional layers of rules and regulations for employees to comply with, making red tape denser at lower levels (cf. Walker and Brewer 2008).

As past research has focused on managers, red tape survey questions often refer to personnel, budget and procurement, or communication activities in ways that are quite specific to managers (e.g., hiring and promoting managers, authorizing expenditures, and counseling low performers). In this sense, these measures fail to capture the full extent of red tape in an organization, especially all of the burdensome and dysfunctional rules that employees encounter in their daily work activities. For instance, employees are often required to perform tasks related to personnel and finance such as reporting personal leave, filling in forms for a training course, or requesting a travel reimbursement. These tasks are related to the management subsystems often examined in red tape research, but in this study, we bore deeper and include the many routine, nonmanagerial tasks that most employees perform, even though these activities are peripheral to their core task. We include personnel and finance activities because they are so prevalent in organizations and because previous studies have focused intently on these managerial activities.

In addition, the job-centered red tape scale examines the rules that pertain to the core task of the employee, which will differ by job type. For employees whose job actually consists of activities in personnel and finance, questions regarding the rules pertaining to their core task may partly overlap with the rules all employees encounter when they complete routine personnel and finance activities, such as reporting personal leave or requesting a travel

reimbursement. However, for employees in other support functions such as information and communications technology (ICT) and for employees whose core task consists of activities in policy development, advice, or implementation, questions about the rules pertaining to their core task probe rules in areas other than routine personnel and finance tasks.

Thus, we develop a measure of red tape that incorporates compliance burden and lack of functionality and taps into what employees know best—the rules they encounter in their daily work activities. In the next section, our research methods and data sources are described.

Methods
This study used a two-step approach in which qualitative and quantitative data were collected to develop and validate a new measure of red tape. Two data sources were used: a relatively small interview group and a larger survey sample.

Data Collection
Step 1. An exploratory qualitative study was conducted to discover whether our theoretical ideas resonated with the perceptions of government employees and to collect input for item development. First, in structured interviews, we asked employees to describe the characteristics of good and bad rules. Second, we explored whether and how rules affect the employees' work in all task areas. Interviews provide rich information on individuals' perceptions and facilitate exploring the characteristics of social phenomena (Boeije 2010).

Ten in-depth interviews were conducted with civil servants from the national government in the Netherlands. Ten interviews are not enough for a complete analysis of red tape perceptions among employees, but that was not the purpose. Our aim was to explore whether employees referred spontaneously to the two characteristics of red tape that are distinguished theoretically (i.e., compliance burden and lack of functionality) and how they described these. The interviewees were selected using a most-dissimilar strategy: the selected respondents differed in their job domain (finance, control, policy) and rank (with and without supervisory responsibilities). Because we were interested in the "lived" experiences of public servants, we probed for more than just the two dimensions of red tape as developed theoretically. The interviews started with an open invitation to describe a "good" and a "bad" rule, regulation, or procedure (as in DeHart-Davis 2009). This was followed up with questions on why the example was a good or a bad rule, inviting the interviewees to reflect on the characteristics of the rule at hand. Based on the interviews, which provided support for the two theoretical dimensions of red tape, survey questions were developed.

Step 2. A survey was conducted to validate these measurement items. For this, we canvassed employees working in Dutch central government ministries such as Internal Affairs or Education, Culture and Science and executive agencies such as the taxing authority, penitentiary facilities, and the Immigration and Naturalization Service. The respondents were members of a voluntary panel of employees who were representative of government employees in their sociodemographic characteristics

> The job-centered red tape scale examines the rules that pertain to the core task of the employee, which will differ by job type.

and willing to participate in short surveys that took no longer than five minutes each. In total, 1,203 panel members participated in the survey, producing a response rate of 47 percent (for more information on the panel, see http://www.internetspiegel.nl).

The respondents formed a highly educated sample: almost 70 percent had a higher professional or university education. On the whole, they had considerable experience working for the government: their average tenure was 21 years. Men formed the majority of the study sample (66 percent). On average, the respondents were 51.3 years old. The core task of most respondents was policy development, implementation, or advice (41 percent). A total of 158 respondents (13 percent) worked in the finance or personnel domain, and 311 (26 percent) worked in a supportive job such as ICT; 239 respondents indicated that they fulfilled a different function.

Data Analysis

All the interviews were recorded and transcribed. They were depersonalized to ensure anonymity by giving each interviewee a reference number (1 to 10) and by referring to them as "he" and "him," although both men and women were interviewed. Each interview was conducted by two researchers, who reflected on the interview afterward to see whether their interpretations differed. This procedure helped ensure interpretative validity (Boeije 2010). The transcripts were read and then reread and coded. The analysis focused on the characteristics that employees mentioned when describing rules as red tape, using the concepts of "compliance burden" and "lack of functionality." Statements that did not fit into these categories were coded as "other characteristics/factors."

By interviewing employees in different jobs and ranks, we were able to identify common elements and tap into generic perceptions of red tape. Specific statements of the interviewees were used to develop the questionnaire items because they effectively summarized recurring content across the interviews.

The survey data were used to test the validity of the measurement scale. The items were first recoded in SPSS so that all items point in the same direction; then the analyses were carried out using Mplus v7.11 (Muthén and Muthén 2013). To validate the scale, we tested red tape as a two-dimensional construct in each of the three types of tasks: core, personnel, and finance. We utilized confirmatory factor analysis for this purpose, as we had clear ideas about the underlying structure of the construct based on theory and empirical evidence (Kline 2010). The items reflected the dimensions, and, as we argued in the theoretical framework, the dimensions together formed red tape, resulting in a second-order formative measure (see Kim 2010 for a similar approach). Here we focus on validating the first-order model.

When testing a measurement instrument's validity, an initial model is often adapted to the data set to increase the fit of the measurement instrument (Byrne 2012). With each adaptation, the model fits the characteristics of the specific data set better, but, at the same time, it could become more difficult to use the model on a different data set (Kline 2010; MacCallum, Roznowski, and Necowitz 1992). Overadaptation means the scale could become too well fitted to the original data set, decreasing its validity with other data sets. Therefore, we used a two-step approach (Anderson and Gerbing 1988). The data were split into a calibration sample (for

adapting the model) and a validation sample (for testing the adapted model's fit). This provided a conservative test of how well the model fit the data (Anderson and Gerbing 1988; MacCallum, Roznowski, and Necowitz 1992).

The measurement items are ordinal in nature (ranging from 1 to 5). Therefore, WLSMV estimation, mean- and variance-adjusted weighted least squares using probit estimates, was used to analyze the models and correct for nonnormality (Kline 2010). In this study, we used three fit indices to assess how well the measurement model fits the data: the comparative fit index (CFI), the Tucker-Lewis index (TLI), and the root mean square error of approximation (RMSEA). For CFI and TLI, cutoff values of .90 indicate an acceptable fit, and .95 or higher indicates good fit. For RMSEA, .10 or lower indicates acceptable fit, while .08 or lower indicates good fit (Hu and Bentler 1999; Kline 2010).

To test whether the new scale fit similarly on both the calibration and validation samples, we used a cross-validation approach. That is, we tested the degree of model similarity between the two samples in steps while increasing the restrictions on the equality of the model. This is similar to testing measurement invariance. We first tested whether the measure had the same dimensions (or factors) and items in the two data sets, followed by intercept-only invariance and scalar invariance, in which the loading of each item on the factor was set at the same value for the two samples (Byrne 2012; MacCallum, Roznowski, and Necowitz 1992; Van der Schoot, Lugtig, and Hox 2012).

We also assessed measurement invariance between the three types of tasks (core, personnel, and finance) because this will show whether the scale measures the same construct for each task. If so, metric invariance should be achieved to enable the comparison of scores between task types (Van der Schoot, Lugtig, and Hox 2012). We compared a model in which the factor loadings were allowed to vary to a model in which the factor loadings were constrained between tasks. For cross-validation and measurement invariance testing, we switched to MLR (robust maximum likelihood) estimation because WLSMV cannot handle missing values for one of the items when cross-validating. CFI is seen as the most reliable measure for cross-validating and invariance testing. If the drop in fit is less than .01, the model is cross-validated (Cheung and Rensvold 2002).

To test whether the items are good indicators of the underlying constructs, convergent validity was assessed by inspecting the factor loadings and the average variance extracted (AVE). An AVE above 50 percent indicates that the items explain a sufficient amount of the variance in the factor (Anderson and Gerbing 1988). The correlations between factors indicate whether the dimensions differed enough to achieve discriminant validity. We also tested whether a model in which all items loaded on one factor fits the data better than the proposed two-dimensional model, which is a more stringent test. The reliability of the instrument was tested by calculating Cronbach's alpha and Raykov's rho, which is an appropriate test when using factor scores based on factor loadings (Bacon, Sauer, and Young 1995).

It is important to test the relationship of a new scale to existing scales that aim to measure the same or similar constructs. If our

measure captures red tape, it should comport with other red tape scales. However, because our measure focuses on red tape pertaining to the employee's job, it should not fully overlap with other scales measuring organizational red tape. To test for this combination of convergent and discriminant validity, a structural equation model was estimated correlating the new scale's dimensions with the GRT scale (in which managers rate the overall level of red tape in their organization on a scale from 1 to 10) and a "compliance time delay" (CTD) scale (measuring the percentage of time respondents say they spend complying with red tape).

Results

Step 1: Qualitative Exploration and Scale Development

The interviewees described examples of good and bad rules encountered in their jobs. Those from different job domains referred to different rules, but even when elaborating on widely different examples, they often gave similar explanations of why a rule was bad. These explanations usually touched on both the compliance burden and lack of functionality of the rule. For example, a director with job responsibilities in the policy domain talked about a procurement rule obliging the department to purchase ICT support from the cheapest bidder:

> It could be that the person leading this procurement says: "this is cheap, having a call center in [place]," but the effect is disastrous. For us it leads to costs and frustration, it misses the goal—in fact, you do not reach your goal at all. (Respondent 3 [hereafter R3]: policy domain, director)

The interviewee referred to the compliance burden as "costs and frustration." He considered that one purpose of the procurement rule is to save money but noted that this could have a disastrous effect (i.e., poor ICT support). He concluded that the rule "misses the goal" and is dysfunctional.

Another example is the description of a bad tendering (or contracting) rule as "not functional, as it does not realize the purpose of the rule" and burdensome because "doing all the work of issuing a tender, selecting a bid, and still not get what you want is frustrating" (R4: finance and control, director). Explaining his perceptions of rules, this interviewee indicated that working for government means you should not only be concerned about the burden in terms of time and cost but also the functionality of a rule when judging its merit. Sometimes "citizens want to know," and this transparency (i.e., functional purpose of the rule) outweighs its burden. He explained,

> We could also change the rule to: when dealing with expenditures below 500 euro we will pay without checking. This would decrease the administrative burden tremendously. But the government cannot do that. The government has to be transparent in everything it does and be accountable. Citizens want to know where every eurocent went. (R4: finance and control, director)

We found that employees connected the compliance burden with lack of functionality in many of their descriptions. The two excerpts that follow show how interviewees referred to both characteristics when explaining why some rules are bad and also how they allowed that compliance burden may be acceptable if a rule serves a useful purpose. These examples come from a supervisor in the finance and control area and an employee in the policy domain who were reflecting on the rules and procedures that pertain to their core activities:

> The fact that you have to undertake a series of actions which apparently have no purpose, that brings about resentment. People working here are usually highly educated and operate more on the basis of thinking for oneself than on instructions. So I think that people will decide: if it serves no purpose, why should I do it? (R9: finance and control, supervisor)

> There are procedures of which you think: what is the purpose of this? There must have been a problem that led to the procedure. You do not see a lot resulting from the time you invest. The added value is often not visible. If it were visible, you would be more willing to comply without complaining. If it is a well-defined procedure you can just work with it. (R1: policy domain, employee)

Some interviewees emphasized a rule's compliance burden, while others emphasized its lack of functionality when explaining why a rule was bad, but both characteristics were usually mentioned. The following example shows how a policy director reflected on the relationship between compliance burden and lack of functionality. The case he was commenting on concerned a personnel rule that gives priority to internal candidates when filling a job vacancy. The purpose of the rule is to ensure fair process, but it entails a disproportionate compliance burden:

> The size of the case should determine how many people have to look at it. Proportionality. The same holds for personnel tasks. When it comes to hiring people, I understand that downsizing the government is not easy when every department keeps hiring from outside while people inside still have to find new positions. Nevertheless, the number of procedures you have to go through [for internal hiring] takes too much time I think. (R8: policy domain, director)

The respondents showed the relevance of both compliance burden and lack of functionality in their perceptions of red tape. The interviews also revealed that these characteristics of red tape were described similarly across all job domains, from policy development to finance and control.

From the interviews we draw support, first, for developing scales that measure each dimension (compliance burden and lack of functionality) separately. Collapsing the two dimensions into a single measure conflates these properties and fails to provide insight into how employees weight the two dimensions when judging whether a rule is good or bad. We measure both dimensions as Bozeman (1993) described them, thus building on and extending past research.

Second, the interviews indicate that employees primarily reflected on the rules that pertain to their own work. A red tape measure that hews closely to the actual experiences of employees should gauge red tape more accurately than measures focusing on the

organization's general rule environment and, more specifically, rules that degrade the organization's performance (Bozeman 2000). Similarly, our job-centered measure should be more accurate than stakeholder red tape, which refers to the way rules affect stakeholder groups rather than individual employees (e.g., Bozeman 2000, 83; Bozeman and Feeney 2011; Feeney and Bozeman 2009). Note that all of these measures are taken from individual reports of red tape, but they differ in what employees were asked to report on—that is, individual, stakeholder group, or organization-wide effects.

Third, to obtain a comprehensive assessment of red tape, questionnaire items should be applicable to employees in different job domains and at different levels of the hierarchy. We met this imperative by asking a robust sample of public servants about their perceptions of rules regarding their "core task." This enabled us to assess, for example, how public servants in the policy domain see rules regarding policy development and advice and how public servants in the finance and control domain see the procurement rules they work with—all by using the same set of survey questions.

Scale development. For each dimension of red tape, several questionnaire items were developed. Statements recorded during the interviews were used as a starting point for developing the items, and they were modified to make them more succinct and broadly applicable when necessary. The interviewees used words such as "time," "frustration," "pressure," "delay," and "compliance" when describing a rule's compliance burden and "goal," "functionality," and "function" when describing its lack of functionality. These words were integrated into the measurement items. Compliance burden was measured with statements on the burden, frustration, and delay caused by complying with the rule. Items regarding the rules' lack of functionality were formulated as statements about its functionality and later reversed. This operationalization is based on Bozeman's (1993) description of lack of functionality as "no efficacy for the rules' functional object."

The scales were pilot tested on a small sample of government officials and academic researchers to refine and improve the measurement items before the full survey was administered. On the basis of their feedback, some unclear and overlapping items were corrected. The items were then checked for consistency, content, and relevance by research officials in the Ministry of Interior Affairs, who supervised the voluntary panel. Some items were slightly revised following their suggestions.

Table 1 Job-Centered Red Tape Measurement Scale

The rules with which I have to comply in my *core activities/personnel activities/ procurement, finance and control activities*:

Lack of functionality
0 have a clear function for my job activities (reversed)
2 contribute to the goal of my job activities (reversed)
4 help me do my job well (reversed)
7 serve a useful goal (reversed)
Compliance burden
1 cause much pressure at work
3 are easy to comply with (reversed)
5 take a lot of time to comply with
6 cause much delay
8 cause a lot of frustration

Table 1 presents the final version of the questionnaire used in the survey. Questions 0, 2, 4, and 7 measure functionality, and questions 1, 3, 5, 6, and 8 measure compliance burden. The item order was thus randomized. All responses were entered on a Likert-type scale ranging from 1 = "totally disagree" to 5 = "totally agree." The items measuring functionality were recoded so that all items point in the same—negative—direction, representing lack of functionality. Because red tape can be present in different types of tasks, all questions were asked three times—first regarding the employee's core task, second regarding personnel-related tasks, and third regarding procurement, finance, and control tasks. The questionnaire introduced the questions with a short clarification. Brief examples were given of some common rules mentioned in the interviews.

Step 2: Validation of the Measure
Descriptive statistics and inter-item correlations. Table A1 in the appendix shows the descriptive statistics and inter-item correlations. All items measuring lack of functionality were positively and moderately or strongly related to each other. The same held for the compliance burden items. Item CB3 was moderately related to most items in both dimensions (above $r = .400$). After reviewing the content of the item, it was excluded because of its conceptual overlap with other items and possible difficulty in interpreting the question.

Results from calibration sample. To develop the model, the sample was split randomly, and the first round of analyses was conducted on the first half of the sample. Model 1, which included two dimensions with four items per dimension, was tested on type of task (core, personnel, and finance) using confirmatory factor analysis (see table 2). For the core and financial tasks, this model fit quite well, but the RMSEA above .10 indicated the models could be improved. For personnel tasks, TLI was below the recommended threshold.

To improve the fit, item 7 was deleted. This item was problematic because it referred to a "useful goal," and respondents seemed reluctant to say that some goals are not useful at all. The model was reestimated for each type of task (model 2). The fit improved,

Table 2 Model Fit Indices for Calibration Sample (LF = lack of functionality, CB = compliance burden)

Model Calibration Sample	CFI	TLI	RMSEA	df	N
Core task					
Model 1: LF = 0-2-4-7; CB = 1-5-6-8	.978	.968	.120	19	597
Model 2: LF = 0-2-4; CB = 1-5-6-8	.980	.968	.199	13	597
Model 3: LF = 0-2-4; CB = 1-5-6	.997	.994	.057	8	597
Personnel					
Model 1: LF = 0-2-4-7; CB = 1-5-6-8	.958	.938	.166	19	591
Model 2: LF = 0-2-4; CB = 1-5-6-8	.985	.976	.117	13	591
Model 3: LF = 0-2-4; CB = 1-5-6	.997	.994	.065	8	591
Finance					
Model 1 LF = 0-2-4-7; CB = 1-5-6-8	.976	.964	.142	19	525
Model 2 LF = 0-2-4; CB = 1-5-6-8	.993	.989	.088	13	524
Model 3 LF = 0-2-4; CB = 1-5-6	.998	.996	.051	8	524

Table 3 Model Fit Indices for Validation Sample (LF=lack of functionality, CB=compliance burden)

Model Validation Sample	CFI	TLI	RMSEA	df	N
Core task					
Model 3 LF=0-2-4; CB=1-5-6	.993	.987	.082	8	598
Personnel					
Model 3 LF=0-2-4; CB=1-5-6	.993	.988	.093	8	594
Finance					
Model 3 LF=0-2-4; CB=1-5-6	.988	.977	.118	8	502

but not sufficiently (RMSEA still too high), and therefore item 8 was also deleted. Item 8 referred to frustration as such, which may be unrelated to compliance burden. Its deletion improved the model substantially and produced a good fit for each type of task (model 3). Factor loadings were all significant and above .700.

Results from validation sample. The final model derived from the foregoing analysis was then tested for each type of task using the second half of the sample (see table 3). This validation analysis shows that the model also fit this sample, although RMSEA was suboptimal for finance tasks. Two of the three fit indices do, however, indicate good fit, so we proceeded with this model.

To see whether the scale measures the construct similarly in the first and second halves of the sample, we conducted a multigroup analysis to cross-validate the model. The results are shown in table A2 in the appendix. When moving from metric invariance, in which the intercepts can differ, to scalar invariance (with factor loadings fixed), the model showed good fit for all tasks. None of the steps indicated a significant decline in CFI (ΔCFI < .01). The results show that the model consisting of the new red tape scale fit equally well across the two samples.

Finally, the measurement invariance between core-task red tape, personnel red tape, and finance red tape was tested to ensure the scale measures the same construct in each task domain. For this, the whole data set was taken. The analysis (see table A3 in the appendix) shows that the items measure the same construct in each type of task because the fit indices remained near the same level. CFI decreased less than the recommended threshold of .01 (ΔCFI = .004).

Validity and reliability. The full data set was used to test the validity and reliability of the scale. The fit and factor loadings of the final model (model 3 with three items per dimension) indicate that this model worked well for each task (see table A4 in the appendix). The AVE for each dimension in the task areas, indicating whether items explain sufficient variance in the underlying construct, ranged from 67 percent to 73 percent, well above the recommended 50 percent threshold. This provides evidence of convergent validity for the scale. To assess reliability, both Raykov's rho and Cronbach's alpha were calculated. Both tests indicate reliability when coefficients are above .70, and the results show that each was well above this threshold (see table A4).

Discriminant validity, showing whether the dimensions are sufficiently different, can be checked by looking at the correlations between the dimensions (Fornell and Larcker 1981). The coefficients were r = .334 between compliance burden and lack of functionality for the core task, r = .186 for personnel tasks, and r = .146 for finance tasks, which shows the overlap was limited. As a more stringent test, the items were modeled to load on one "red tape" construct per task. For all three task types, this one-dimensional model fit poorly, with all fit indices well below the recommended thresholds (core task: CFI = .808, TLI = .679, RMSEA = .398, personnel tasks: CFI = .802, TLI = .669, RMSEA = .473, finance tasks: CFI = .772, TLI = .619, RMSEA = .503).

To assess discriminant validity regarding other red tape measures, we tested our measure's relationship to the GRT scale and the CTD measure, which were also collected in the survey. Our measure should be related to existing measures because all are attempting to measure the same underlying concept (red tape); however, our measure should also be sufficiently different from existing measures to show that it is not redundant but rather provides a fuller picture of red tape.

Figure 2 shows the structural equation model in which the GRT and CTD measures were correlated with the new measure's two dimensions of red tape in the core task of the respondents. The GRT was significantly related to both dimensions of the new scale but most strongly to the compliance burden. The same held for the CTD measure. The GRT and CTD measures were more strongly correlated with each other (.596) than with the new measure. This
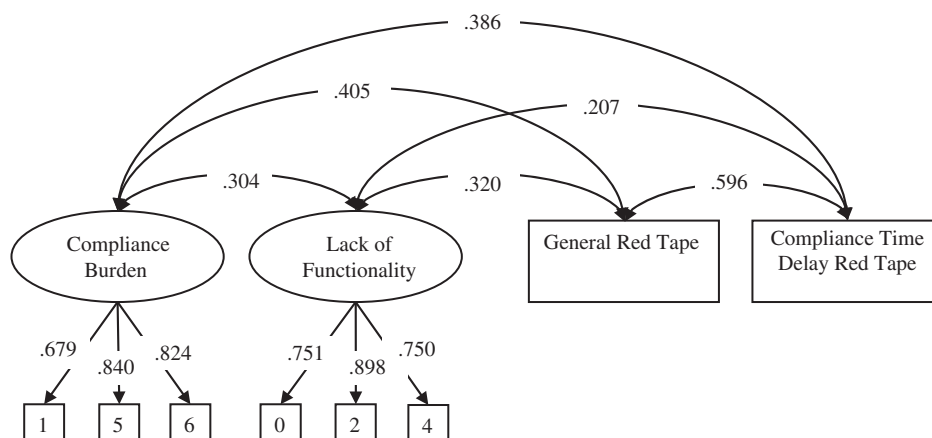

Figure 2 Correlation between Measures, Standardized Estimates (CFI = .980, TLI = .965, RMSEA = .055, SRMR = .029)

Lack of functionality

↑

High

Unnecessary rules	Red Tape
Core: 7.5%	Core: 20.3%
Personnel: 27.4%	Personnel: 38.9%
Finance: 18.7%	Finance: 39.3%

→ Compliance burden

Low   High

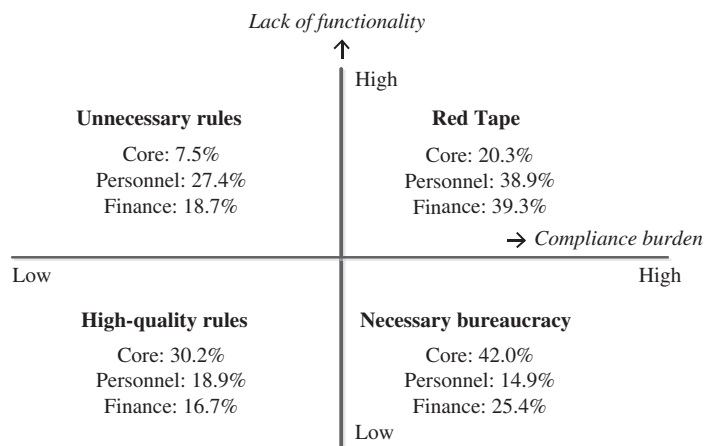| High-quality rules	Necessary bureaucracy
Core: 30.2%	Core: 42.0%
Personnel: 18.9%	Personnel: 14.9%
Finance: 16.7%	Finance: 25.4%

Low

Figure 3 Percentage of Respondents Classifying Rules (for Core, Personnel, and Finance Tasks) Based on the Compliance Burden and Lack of Functionality of the Rules

analysis reveals that the new scale is significantly related to the existing scales but also substantially different. In a final test (not reported), we simultaneously regressed the GRT and our new scale on self-reported work quality. This step showed that both scales were negatively and significantly related to perceptions of work quality, but when both dimensions were included simultaneously, only the lack of functionality dimension of our measure was significantly related. This finding further confirms that our new measure taps into the neglected dimension of red tape known as rule functionality. We acknowledge that self-reported work quality may be subject to common source bias; therefore, further research is needed to establish definitive evidence of the new scale's added value.

Employees' Job-Centered Rule Perceptions
The validated scale of job-centered red tape allows us to examine how employees perceive the rules pertaining to their jobs. Depending on the scores for the two rule characteristics—compliance burden and lack of functionality—we can classify rules for different types of tasks into the following taxonomy: "red tape," "necessary bureaucracy," "unnecessary rules," and "high-quality rules." For the classification, we first calculated the dimensional score by taking the mean of the items comprising the scale on each type of task (core, personnel, and finance). Next, we designated the middle of the scale (3.0) as the cutoff point. Mean scores *above* the cutoff indicate a lack of functionality and a high compliance burden.

Figure 3 summarizes the results. For the employee's core task, 20.3 percent of the respondents perceived the rules as red tape (i.e., both as highly burdensome and lacking functionality). For respondents' personnel and finance tasks, however, roughly twice as many perceived the rules as red tape (38.9 percent and 39.3 percent, respectively).

There are many interesting findings in this breakdown of rule characteristics, including the fact that respondents on the whole reacted more favorably to rules involving their core task, which they presumably encountered more frequently than other types of rules. In addition, among the rules encountered less frequently, personnel rules were seen as less necessary on the whole than finance rules.

> Red tape affecting an employee's core task may have a larger and more negative impact on the organization, given the centrality and importance of line tasks compared with staff tasks.

Discussion
Overall, this study provides support for an approach to measuring red tape that explicitly focuses on the two characteristics of rules that constitute red tape—compliance burden and lack of functionality—and concentrates on how employees experience rules in their work rather than in more abstract, remote ways such as organizational or stakeholder group red tape. Our analyses of the validity and reliability of the newly developed scale show that the scale is a valuable supplement to existing scales.

The finding that most widely used scales tap the compliance burden rather than the lack of functionality dimension of red tape raises questions about previous studies using those scales. As Feeney (2012) explains, the GRT may not entirely capture the fact that some rules, although burdensome, have an important purpose. Here, using a measure that includes lack of functionality as one of the determining factors, we find that 20.3 percent of the employees perceived red tape in their core task area. This finding raises questions about the magnitude and severity of the red tape problem overall. Because other red tape scales mostly measure compliance burden and, to a lesser degree, lack of functionality, researchers should be careful when drawing conclusions about the amount of red tape in public organizations and how it can be managed.

This study confirms that red tape exists and that it hinders employees when the rule's compliance burden outweighs its functionality or purpose. Addressing these more extreme cases can be an important way to improve performance of public services. One implication is that compliance burden and functionality should be properly balanced for rules to work well. If the compliance burden outweighs the rule's functional value, managers should try to restore the balance by either reducing the compliance burden when possible or clarifying the rule's functional purpose, which may be unclear or misunderstood by employees. The latter could well be related to the tendency in organizations to impose rules without briefing employees on their purpose, and this can lead to misperceptions about the purpose and value of the rules. In rare cases of a nonfunctional or dysfunctional rule, managers should explain the rule's detrimental effect to rulemaking authorities and advocate for its removal.

Interestingly, the final analyses show that considerably more employees perceived rules regarding personnel and finance activities as red tape compared with rules in their core task area. This result confirms what Coursey and Pandey (2007) and others have reported: employees perceive more red tape in personnel and finance activities than in other areas. This difference in red tape perceptions across tasks may be attributable to employees having a better understanding of the purpose and function of rules in their core tasks compared with personnel and financial rules, which may be more peripheral for the employee. However, red tape affecting an employee's core task may have a larger and more negative impact on the organization, given the centrality and importance of line tasks compared with staff tasks. This is an interesting issue for further research.

This study is a first step in developing and validating a new job-centered red tape

scale. To validate the scale further, this study should be replicated, preferably in a different context so as to test its general applicability. Researchers could survey other types of public employees, in different public domains, or in different countries. Moreover, one feature of this scale is that the dimensions are measured in different directions. This is how the interviewees make sense of these dimensions in what effectively are high-quality rules (see questionnaire items in table 1). Yet this difference could provide an alternative explanation for the two-factor solution, even though the items on functionality were reversed. In addition, we have attempted to validate the first-order construct of red tape in this study. Future research needs to explore whether the dimensions are formative or reflective. We argue that the dimensions are formative in that, together, they form red tape perceptions. Coursey and Pandey (2007) found evidence of red tape as reflective construct but actually focused on various management subsystems of red tape, which is a different issue. Future research could, for instance, examine our measure's relationship with relevant outcome variables such as job performance and organizational effectiveness. In doing so, it can test whether our measure is a second order construct, whether there are two separate dimensions, and whether the dimensions interact. Finally, two items were deleted from our scale to produce acceptable fit, leaving three items per dimension. While this is a sufficient number of items going forward, researchers may want to continue searching for additional measurement items.

This scale can be utilized by researchers and practitioners alike. First, it can be used for a fine-grained understanding of red tape. Because the two dimensions of red tape are distinct, for example, applying the scale makes it possible to study whether the compliance burden or lack of functionality of rules most affects employees' job satisfaction and performance. Brewer and Walker (2010b) found that various types of red tape are related to various dimensions of performance but in different and sometimes unexpected ways. This is of interest because one dimension of red tape may have more negative consequences than the other and thus warrant more attention, both in rulemaking and rule enforcement. Second, this scale can be used to explore whether there is a trade-off between the two dimensions and to what extent employees are tolerant of burdensome but necessary rules. A longitudinal design could also provide insight into how rules can be improved to decrease their compliance burden and enhance their perceived functionality.

Because the scale is relatively short and stable across multiple job activities, managers can use the instrument for a "quick scan" of the organization to identify which units are experiencing high levels of red tape and plan interventions accordingly. The new red tape scale makes it possible to compare red tape across organizational units and zoom in on departments or teams. Unit managers can then use this information as a starting point for discussions with employees about the rules in force and how these rules might be improved. Such practical applications could result in less red tape, higher levels of job satisfaction, smoother service delivery, and improved public service.

Conclusion
The objective of this study was to develop a measure of red tape that addresses three measurement concerns of existing scales. First, we argued that existing measures tend to focus explicitly on

the compliance burden aspect of red tape while neglecting rule functionality. In our interviews, individuals distinguished between these dimensions and used them in tandem to describe good rules and bad rules. Second, most existing measures ask employees to report on organizational or stakeholder red tape, both of which encompass rules that may be distant to employees or lie beyond their knowledge and experience. Focusing on rules that pertain to employees' job activities is thus a major improvement. The ensuing statistical analysis confirmed that compliance burden and lack of functionality are related but distinct characteristics of rules and presented evidence that the more neglected dimension of rule functionality may be the most important characteristic. The analyses of the survey data provided evidence for the validity of our job-centered red tape scale. Third, for a more complete and accurate view of red tape, we focused on the core activities that employees perform in their jobs and other collateral duties that may be required, such as complying with personnel and financial rules. By focusing on rules that employees have firsthand knowledge of, and by explicitly distinguishing between core task rules and personnel and finance related rules, we were able to draw some useful insights about the rule environment of public organizations.

The job-centered red tape scale developed here is meant to supplement existing scales. It can be applied across public organizations to examine how employees assess the compliance burden and lack of functionality of the rules encountered in their work. This measure can also be used to compare departments and teams within an organization so that managers can identify problem areas and plan interventions. Overall, our new job-centered red tape scale may prove valuable to researchers and practitioners who are trying to understand the elusive nature of red tape.

Acknowledgments

References
Anderson, James C., and David W. Gerbing. 1988. Structural Equation Modeling in Practice: A Review and Recommended Two-Step Approach. *Psychological Bulletin* 103(3): 411–23.

Bacon, Donald R., Paul L. Sauer, and Murray Young. 1995. Composite Reliability in Structural Equations Modeling. *Educational and Psychological Measurement* 55(3): 394–406.

Boeije, Hennie. 2010. *Analysis in Qualitative Research*. Los Angeles: Sage Publications.

Bozeman, Barry. 1993. A Theory of Government "Red Tape." *Journal of Public Administration Research and Theory* 3(3): 273–303.

———. 2000. *Bureaucracy and Red Tape*. Upper Saddle River, NJ: Prentice Hall.

———. 2012. Multidimensional Red Tape: A Theory Coda. *International Public Management Journal* 15(3): 245–65.

Bozeman, Barry, and Mary K. Feeney. 2011. *Rules and Red Tape: A Prism for Public Administration Theory and Research*. Armonk, NY: M. E. Sharpe.

Brewer, Gene A. 2013. Public Management Contributions for Improving Social Service Performance: Public Values, Public Service Motivation, and Rule Functionality. In *Managing Social Issues: A Public Values Perspective*, edited by

Peter L. M. Leisink, Paul Boselie, Maarten van Bottenburg, and Dian Marie Hosking, 19–36. Northampton, MA: Edward Elgar.

Brewer, Gene A., and Richard M. Walker. 2010a. Explaining Variation in Perceptions of Red Tape: A Professionalism-Marketization Model. *Public Administration* 88(2): 418–38.

———. 2010b. The Impact of Red Tape on Governmental Performance: An Empirical Analysis. *Journal of Public Administration Research and Theory* 20(1): 233–57.

———. 2012. Red Tape: The Bane of Public Organizations? In *Public Management and Performance: Research Directions,* edited by Richard M. Walker, George A. Boyne, and Gene A. Brewer, 110–26. New York: Cambridge University Press.

Brewer, Gene A., Richard M. Walker, Barry Bozeman, Claudia N. Avellaneda, and Gene A. Brewer, Jr. 2012. External Control and Red Tape: The Mediating Effects of Client and Organizational Feedback. *International Public Management Journal* 15(3): 288–314.

Byrne, Barbara M. 2012. *Structural Equation Modeling with Mplus: Basic Concepts, Applications, and Programming.* New York: Routledge Academic.

Chen, Greg, and Daniel Williams. 2007. How Political Support Influences Red Tape through Developmental Culture. *Policy Studies Journal* 35(3): 419–36.

Cheung, Gordon W., and Roger B. Rensvold. 2002. Evaluating Goodness-of-Fit Indexes for Testing Measurement Invariance. *Structural Equation Modeling: A Multidisciplinary Journal* 9(2): 233–55.

Coursey, David H., and Sanjay K. Pandey. 2007. Content Domain, Measurement, and Validity of the Red Tape Construct: A Second-Order Confirmatory Factor Analysis. *American Review of Public Administration* 37(3): 342–61.

DeHart-Davis, Leisha. 2009. Green Tape: A Theory of Effective Organizational Rules. *Journal of Public Administration Research and Theory* 19(2): 361–84.

DeHart-Davis, Leisha, and Sanjay K. Pandey. 2005. Red Tape and Public Employees: Does Perceived Rule Dysfunction Alienate Managers? *Journal of Public Administration Research and Theory* 15(1): 133–48.

DeVellis, Robert F. 2003. *Scale Development: Theory and Applications.* Thousand Oaks, CA: Sage Publications.

Feeney, Mary K. 2012. Organizational Red Tape: A Measurement Experiment. *Journal of Public Administration Research and Theory* 22(3): 427–44.

Feeney, Mary K., and Barry Bozeman. 2009. Stakeholder Red Tape: Comparing Perceptions of Public Managers and Their Private Consultants. *Public Administration Review* 69(4): 710–26.

Fornell, Claes, and David F. Larcker. 1981. Evaluating Structural Equation Models with Unobservable Variables and Measurement Error. *Journal of Marketing Research* 18(1): 39–50.

Frederickson, H. George. 2005. The State of Social Equity in American Public Administration. *National Civic Review* 94(4): 31–38.

Hu, Li-tze, and Peter M. Bentler. 1999. Cutoff Criteria for Fit Indexes in Covariance Structure Analysis: Conventional Criteria versus New Alternatives. *Structural Equation Modeling: A Multidisciplinary Journal* 6(1): 1–55.

Kaufman, Herbert. 1977. *Red Tape: Its Origins, Uses, and Abuses.* Washington, DC: Brookings Institution.

Kim, Sangmook. 2010. Testing a Revised Measure of Public Service Motivation: Reflective versus Formative Specification. *Journal of Public Administration Research and Theory* 21(3): 521–46.

Kline, Rex B. 2010. *Principles and Practice of Structural Equation Modeling*, 3rd ed. New York: Guilford Press.

MacCallum, Robert C., Mary Roznowski, and Lawrence B. Necowitz. 1992. Model Modifications in Covariance Structure Analysis: The Problem of Capitalization on Chance. *Psychological Bulletin* 111(3): 490–504.

Moynihan, Donald P., and Sanjay K. Pandey. 2007. The Role of Organizations in Fostering Public Service Motivation. *Public Administration Review* 67(1): 40–53.

Moynihan, Donald P., Bradley E. Wright, and Sanjay K. Pandey. 2012. Working within Constraints: Can Transformational Leaders Alter the Experience of Red Tape? *International Public Management Journal* 15(3): 315–36.

Muthén, Linda K., and Bengt O. Muthén, eds. 2013. *Mplus User's Guide*, 7th ed. Los Angeles: Muthén & Muthén.

Pandey, Sanjay K., and Stuart I. Bretschneider. 1997. The Impact of Red Tape's Administrative Delay on Organizations' Interest in New Information Technologies. *Journal of Public Administration Research and Theory* 7(1): 113–30.

Pandey, Sanjay K., David H. Coursey, and Donald P. Moynihan. 2007. Organizational Effectiveness and Bureaucratic Red Tape: A Multi-Method Study. *Public Performance and Management Review* 30(3): 398–425.

Pandey, Sanjay K., and Gorden A. Kingsley. 2000. Examining Red Tape in Public and Private Organizations: Alternative Explanations from a Social Psychological Model. *Journal of Public Administration Research and Theory* 10(4): 779–800.

Pandey, Sanjay K., and Patrick G. Scott. 2002. Red Tape: A Review and Assessment of Concepts and Measures. *Journal of Public Administration Research and Theory* 12(4): 553–80.

Quratulain, Samina, and Abdul K. Khan. 2015. Red Tape, Resigned Satisfaction, Public Service Motivation, and Negative Employee Attitudes and Behaviors: Testing a Model of Moderated Mediation. *Review of Public Personnel Administration* 35(4): 307–22.

Rainey, Hal G., Sanjay K. Pandey, and Barry Bozeman. 1995. Public and Private Managers' Perceptions of Red Tape. *Public Administration Review* 55(6): 567–74.

Scott, Patrick G., and Sanjay K. Pandey. 2000. The Influence of Red Tape on Bureaucratic Behavior: An Experimental Simulation. *Journal of Policy Analysis and Management* 19(4): 615–33.

———. 2005. Red Tape and Public Service Motivation: Findings from a National Survey of Managers in Health and Human Services Agencies. *Review of Public Personnel Administration* 25(2): 155–80.

Van der Schoot, Rens, Peter Lugtig, and Joop Hox. 2012. A Checklist for Testing Measurement Invariance. *European Journal of Developmental Psychology* 9(4): 486–92.

Walker, Richard M., and Gene A. Brewer. 2008. An Organizational Echelon Analysis of the Determinants of Red Tape in Public Organizations. *Public Administration Review* 68(6): 1112–27.

Table A1 Items: means and correlations items, $N = 1,189$

	Mean (SD)	LF 0R	LF 2R	LF 4R	LF 7R	CB 1	CB 3R	CB 5	CB 6	CB 8
LF 0R	2.09 (.971)	1								
LF 2R	2.35 (.985)	.676*	1							
LF 4R	2.68 (1.08)	.554*	.675*	1						
LF 7R	2.41 (1.00)	.643*	.712*	.615*	1					
CB 1	3.19 (1.12)	.171*	.215*	.223*	.243*	1				
CB 3R	2.71 (1.01)	.338*	.409*	.439*	.401*	.464*	1			
CB 5	3.11 (1.13)	.152*	.174*	.151*	.209*	.584*	.403*	1		
CB 6	3.07 (1.12)	.238*	.266*	.214*	.291*	.523*	.354*	.700*	1	
CB 8	2.92 (1.15)	.329*	.347*	.329*	.389*	.559*	.518*	.542*	.592*	1

*Correlation is significant at the 0.001 level (two-tailed). LF=lack of functionality, CB=compliance burden.

Table A2 Cross-Validation from Calibration to Validation Sample

Cross-Validation	CFI	ΔCFI	TLI	RMSEA	N
Core task					
1 Metric	.986	—	.981	.046	1,195
2 Intercepts only	.986	.000	.981	.047	1,195
3 Scalar	.987	.001	.986	.041	1,195
Personnel					
1 Metric	.991	—	.988	.038	1,185
2 Intercepts only	.991	.000	.988	.037	1,185
3 Scalar	.991	.000	.990	.034	1,185
Finance					
1 Metric	.986	—	.982	.045	1,026
2 Intercepts only	.987	.001	.982	.044	1,026
3 Scalar	.985	−.002	.984	.041	1,026

Table A3 Measurement Invariance between Types of Tasks

Model	CFI	TLI	RMSEA
1 Factor loadings free	.929	.918	.054 (.051-.058)
2 Factor loadings constrained	.925	.917	.055 (.052-.058)

Table A4 Factor Loadings (standardized), Fit (CFI, TLI, and RMSEA), Validity, and Reliability (Raykov's rho, Cronbach's alpha) of Red Tape Measure (C=core, P=personnel, F=finance)

The rules with which I have to comply in core/personnel/finance activities…	Factor Loadings			Mean			SE			R2		
Lack of functionality	C	P	F	C	P	F	C	P	F	C	P	F
0 have a clear function for my job activities (reversed)	.797	.762	.790	2.09	2.79	2.74	.014	.014	.013	.636	.581	.625
2 contribute to the goal of my job activities (reversed)	.938	.922	.914	2.35	3.39	3.17	.009	.011	.011	.879	.850	.836
4 help me do my job well (reversed)	.785	.838	.851	2.68	3.52	3.29	.014	.012	.012	.617	.702	.724
AVE	.71	.71	.73									
Raykov's rho	.91	.90	.90									
Cronbach's alpha	.84	.82	.86									
Compliance burden	C	P	F	C	P	F	C	P	F	C	P	F
1 cause much pressure at work	.708	.740	.788	3.19	2.93	3.09	.016	.013	.012	.501	.547	.621
5 take a lot of time to comply with	.877	.892	.910	3.11	2.99	3.22	.011	.009	.009	.769	.796	.827
6 cause much delay	.857	.902	.864	3.07	2.95	3.27	.012	.009	.010	.735	.814	.746
AVE	.67	.72	.73									
Raykov's rho	.88	.90	.90									
Cronbach's alpha	.82	.85	.86									
CFI	.995	.997	.994									
TLI	.990	.995	.989									
RMSEA	.070	.058	.086									
N	1,195	1,185	1,026									