Measuring red tape in a hospital setting: A survey experiment

Peer-reviewed author version

LUYTEN, Janis & MARNEFFE, Wim (2022) Measuring red tape in a hospital setting: A survey experiment. In: International review of administrative sciences, 89 (4) , p. 958-976.

DOI: 10.1177/00208523211073498
Handle: http://hdl.handle.net/1942/36652

Measuring Red Tape in a Hospital Setting: a Survey Experiment

Janis Luyten and Wim Marneffe

Janis Luyten

First author

janis.luyten@uhasselt.be

Faculty of Business Economics, Hasselt University

Martelarenlaan 42, 3500 Hasselt

BELGIUM

Wim Marneffe

Corresponding author

wim.marneffe@uhasselt.be

Faculty of Business Economics, Hasselt University

Martelarenlaan 42, 3500 Hasselt

BELGIUM

Abstract

Public administration research is actively exploring alternatives for the General Red Tape (GRT) scale to measure red tape. Due to increasing criticism on the GRT scale, scholars proposed the Three-Item Red Tape (TIRT) scale as an alternative. Using a repeated cross-sectional design, this article tests both scales in a before-after analysis of a major change in the organization of administration in a hospital. The results indicate that the GRT scale does not capture the resulting major change in red tape, which raises questions on the instrument's validity in a bottom-up research design within one organization. The TIRT scale, however, which measures red tape at the work environment level, does reflect the change in red tape but shows empirical weaknesses in its design. Additionally, by randomly assigning respondents to

substantially different red tape definitions, this article shows that the red tape definition does not significantly impact respondents' GRT ratings.

Points for practitioners

- The predominantly used General Red Tape scale is not able to capture an increase in red tape in a bottom-up intraorganizational research design in a hospital, which raises questions on the instrument's validity.
- A more recent alternative for the General Red Tape scale, more specifically the Three-Item Red Tape scale, captures the increase in red tape but shows empirical weaknesses.
- The wording of the red tape definition does not impact respondents' answers on the General Red Tape scale.

Keywords: Red tape , General Red Tape scale , Three-item Red Tape scale , survey experiment , Public Administration, hospital.

Authors short Bio

Dr. Janis Luyten, PhD in Business Economics. Faculty of Business Economics, Hasselt University, Belgium.

Research fields: public administration, health economics, public economics.

Prof. Dr. Wim Marneffe, Professor Business Economics. Faculty of Business Economics, Hasselt University, Belgium.

Research fields: public administration, law and economics, health economics, public economics.

Introduction

Recent empirical public administration research focuses mainly on the effects of red tape on organizational or individual factors and the drivers of red tape perceptions (see e.g. Migchelbrink and Van de Walle (2020), Oh and Lee (2020), Kaufmann and Tummers (2017), Jacobsen and Jakobsen (2018)). Scholars have demonstrated that reducing red tape leads to enhanced organizational performance, which emphasizes the importance of the concept in public and private policy-making (see e.g. De Jong and Van Witteloostuijn (2015), Walker and Brewer (2009)). To date, however, significant uncertainty remains in the literature on how to measure red tape. This article contributes to the existing red tape literature by examining two major methodological shortcomings using an experimental survey design.

First, most scholars employ the General Red Tape (GRT) scale, introduced by Rainey et al. (1995), to gain insight into red tape. Over the years, however, the GRT scale has received a substantial amount of criticism (see e.g. Luton (2007), Bozeman and Feeney (2011)), which resulted in scholars shaping new measures, including the Three-Item Red Tape (TIRT) scale of Borry (2016). One of the critiques is aimed at the definition used in the GRT scale. Scholars claim the definition carries a negative connotation, triggering respondents to give a more negative response on the scale (see e.g. Bozeman and Feeney (2011), Feeney (2012)). Moreover, Feeney (2012) finds ratings of perceived organizational red tape to be substantially influenced by the definition accompanying the GRT scale. Another frequently used red tape definition is the European Commission's (EC) conceptualization of red tape, which is considered to be more practice-based, and distinguishes rules from red tape more explicitly compared to the original GRT definition (European Commission, 2017). Using unique, primary survey data from a hospital and by randomly assigning one of the two definitions to respondents, this article tests whether the random presentation of the original GRT definition or the EC definition leads to variations in red tape ratings.

Second, to date, the predominant research approach of red tape scholars is to collect data from multiple organizations each represented by one manager to gain insight into organizational red tape at a given point in time, using a top-down oriented cross-sectional research design. Recently, however, leading red tape scholars identified this research approach as one of the caveats currently present in red tape research (see e.g. Kaufmann et al. (2020)). The predominantly used cross-sectional design does not allow researchers to assess the validity of red tape instruments. A repeated cross-sectional design, however, does provide the possibility to examine the validity of the instrument, especially when a fundamental transformation of the administrative processes changes red tape levels, by testing whether this change is reflected in the instrument's results. In this study, the major transformation in administrative processes was the implementation of an integrated Electronic Health Records (EHR) system, which replaced various fragmented software packages on a single day, and therefore provides a clear cut-off in time for the analyses. Hence, the second aim of this article is to examine two red tape instruments (i.e. GRT and TIRT) over time and their sensitivity to major changes in red tape using a bottom-up oriented, repeated cross-sectional research design in a single organization.

Theory

In his theoretical contribution to red tape research, Bozeman (1993) conceptualized red tape as "rules, regulations, and procedures that remain in force and entail a compliance burden for the organization but have no efficacy for a rule's functional object". To examine the differences in perceived red tape between private and public managers, Rainey et al. (1995) used this definition to shape the General Red Tape (GRT) scale. The GRT measure is a scale based answer to the question "If red tape is defined as burdensome rules and procedures that have negative effects on the organization's effectiveness, how would you assess the level of red tape in your organization?". Respondents are asked to indicate the level of perceived organizational red tape on a 0 to 10 scale in which 0 corresponds to almost no red tape and 10 represents a great deal of red tape, with the purpose of gaining insight into perceived levels of organizational red tape.

Since its introduction, mostly due to its ease of application and generic nature, the GRT scale has been used in a wide variety of domains. Bozeman and Feeney (2011) argue the GRT scale is one of the best examples of consistent, cross-research application of an empirical measure in public administration research. The scale has been applied in surveys administered to public, private and non-profit managers, health and human services departments, city managers, etc. Yet, empirical red tape research in healthcare organizations is to date fairly limited (Henderson and Borry, 2020; Van Loon, 2017), especially in hospital settings.

Over the years, scholars started criticizing the GRT scale. Luton (2007) argues that respondents first have to decide which rules are burdensome and subsequently if those burdensome rules affect organizational effectiveness in a negative way. Respondents thus have to make two judgments before making the requested evaluation on a single indicator

scale. From the work of Bozeman (Bozeman, 2000; Bozeman and Feeney, 2011), Borry (2016) argues that only rules which are burdensome, unnecessary, and ineffective are considered red tape and that the single indicator GRT scale may not be able to capture this complexity of the red tape concept. In a critical article on red tape research, Bozeman (2012) stated that scholars have chosen simplicity at the cost of realism.

Borry (2016) therefore argues there is a need for a multiple indicator scale to counter previously mentioned critiques. She proposes the Three-Item Red Tape (TIRT) scale, which is composed of three rule characteristics distinguishing rules from red tape: burden, effectiveness, and necessity. The TIRT scale is a multiple indicator scale asking respondents to answer the question "How would you describe policies and procedures in your work division between the following opposite characteristics?" These characteristics range from not burdensome to burdensome, necessary to unnecessary, and effective to ineffective.

The TIRT scale has several advantages over the GRT scale. First, because of the multiple indicator scale scholars gain more insight into the extent that rules are considered to be red tape. This distinction can be of fundamental importance in healthcare. Although rules and procedures may be burdensome, they can be necessary in safeguarding patient safety and quality of care (Van Loon, 2017). Second, the TIRT scale asks respondents to evaluate red tape in their specific work environment in contrast to the entire organization, which is in line with the recent evolution towards multiple stakeholder measurement. Although this could lead to significant differences between GRT and TIRT outcomes, recent articles have found TIRT levels to be very close to GRT levels but weak internal consistency of the TIRT scale (Kaufmann et al., 2020, 2021). The conceptual similarity between the GRT and TIRT scales could be explained by the fact that the TIRT scale was built on the same theoretical foundations as the GRT scale. It is because of this similarity we preferred the TIRT scale over

other recently introduced red tape measures (such as the Job-Centered Measure of Van Loon et al. (2016)).

The aforementioned critiques of leading red tape scholars lead to uncertainty among researchers whether the GRT scale is a valid instrument to measure red tape, and whether the TIRT scale is a viable alternative (Kaufmann et al., 2019). Since the dominant approach in red tape research is to collect red tape ratings of multiple organizations each represented by a single manager at a given point in time, it is difficult to assess the validity of the GRT scale. When multiple measurements, preferably among multiple stakeholders within each organization, would be performed in the same organization(s) over time one could assess the validity of the GRT measurement by manipulating the amount of red tape experienced by respondents and analyse whether this change is reflected in GRT scale's ratings (Pandey and Marlowe, 2014). The aim of this article is to simultaneously investigate the validity of the GRT and TIRT scales using a before-after analysis of a major change in the administrative processes of a hospital in a repeated cross-sectional design. In contrast to Pandey and Marlowe (2014), who changed red tape levels artificially using a vignette-approach in an experimental setting, this article examines a real-life change in red tape due to the transformation of the administrative processes of a hospital avoiding potential hypotheticality bias. This major change was the replacement of various software packages with an integrated Electronic Health Records (EHR) system. The fragmented software packages mainly consisted of free text entry spaces while the new EHR used check boxes in order to promote structured data gathering. These check boxes focused on a wide range of potential risk factors resulting in more extensive reporting compared to the free text entry in which only the most relevant findings were reported. Furthermore, the new EHR consisted of numerous features the previous software packages did not (e.g. calculation of early warning scores), resulting in more extensive reporting. Therefore, based on the abovementioned information, workshops

with hospital staff, and observations at various departments in the hospital both before and after the implementation of the EHR, we argue red tape levels will increase in the hospital and, consequently, red tape ratings on the GRT and TIRT scales are expected to increase as well. This leads to the following hypothesis:

H_1: A major increase in red tape leads to an increase in GRT and TIRT ratings.

Opposite to the vast body of red tape research, mainly using the original GRT scale of Rainey et al. (1995), the European Commission (EC) employs a different approach to measure red tape in the Standard Cost Model. The model is based on two types of administrative costs: the business-as-usual costs and the administrative burdens[1]. Business-as-usual costs are the costs resulting from and time spent on information obligations which would also be done in the absence of procedures and legislation. In other words, these costs are the costs of necessary and effective rules and should therefore not be considered as red tape. The administrative burdens on the other hand are the costs resulting from activities which are done solely because of a legal obligation, and are therefore deemed unnecessary, ineffective, and burdensome (European Commission, 2017). The EC thus uses a substantially different definition to define red tape in contrast to the majority of red tape scholars: "Administrative burdens represent the time and costs incurred in meeting legal obligations to provide information on your activities, either to public or private authorities, which are done solely because of this legal obligation" (European Commission, 2017). The EC conceptualization of red tape is widely used and accepted among policy makers and has been taken over by EU member states, the OECD, and scholars (e.g. Poel et al. (2014), Arendsen et al. (2014), OECD (2007)).

[1] In Dutch the translation of 'red tape' and 'administrative burdens' is the same. Therefore, a distinction between the concepts is not relevant in this context.

In a measurement experiment using three alternative definitions, Feeney (2012) concludes definition wording is of significant importance when using the GRT scale as it affects respondents' evaluations. This experiment distinguishes itself from Feeney's research in three ways. First, this study is performed in a hospital context rather than in government agencies. While local government managers' daily business is administration, hospital staff is mainly focused on the provision of health care. Second, this study uses a bottom-up in contrast to a top-down approach by including multiple hospital staff members with different job contents within one organization. Third, while Feeney used three theoretical red tape definitions as alternatives for the original GRT definition, this study employs a practice-based alternative to define red tape. By randomly assigning respondents to either the GRT or the EC definition, this article examines whether a perceptual difference can be observed between the two alternatives with the final hypothesis:

H_2: Variations in perceived levels of organizational red tape stem from a difference between the GRT and EC definition.

Methods

Study and Survey Design

This article employs primary survey data collected in a Belgian general hospital with five campuses. The hospital accommodates approximately 3,000 health professionals and accounts for more than 650,000 patient consultations each year. This study uses a repeated cross-sectional design to test the aforementioned hypotheses. The reason to opt for multiple measurements in time is to be able to accurately assess the sensitivity of the GRT and TIRT scales to changes in red tape. The hospital participating in this study was to undergo a major change in its administrative processes. This major change was the implementation of an integrated Electronic Health Records system (EHR), which had to replace multiple software packages. The hospital used a rather unique implementation strategy in which the transition from the fragmented software to the integrated EHR took place in a single day, whereas most hospitals opt for gradual implementations of the different modules over time.

A red tape survey was administered at three points in time to conduct a before-after analysis. The first measurement was conducted one month before the implementation of the EHR to avoid other major organizational changes to bias the results (e.g. different work protocols, staffing, etc.). Following the healthcare technology impact and acceptance literature, the second measurement took place 11 months after implementation for two reasons. First, by waiting almost a year we were able to avoid measuring the irritation inherent to a substantial organizational change which could bias red tape perceptions. Second, after using the system for one year, learning effects and the inherent dip in performance should be minimal as hospital staff has gained sufficient experience with the new system. The final measurement was conducted 14 months after implementation. The relatively short time span between the second and third survey was chosen on purpose since the final survey served

as a robustness check for the results obtained in the second survey. Since no changes in organizational processes were carried out during this period, the results of the second and third survey should thus be highly comparable.

The survey was administered to all types of health professionals and generally consisted of three sections: professional profile, red tape perceptions, and respondent characteristics. The professional profile section asked respondents to indicate their profession, experience on the job, amount of hours worked in a week, and campus. The red tape section focussed on the amount of time spent on respectively administration, communication with colleagues, and patient interaction. Furthermore, this section included the GRT and TIRT scales. A detailed overview of the red tape sections in the three surveys is shown in Figures 1-3. In the first and second survey, each respondent was randomly assigned to either the original GRT scale or the EC GRT scale by the survey software package. In order to compare the TIRT and GRT scales, this article uses a 0-10 scale for the TIRT scale instead of the 5-point scale proposed by Borry (2016). Note that in the surveys containing both the GRT and TIRT scales, the TIRT scale, which was identical to all respondents, precedes the GRT scale with alternating definitions to avoid biased TIRT ratings by the definition respondents were presented when completing the GRT scale. Because of this reason and to achieve sufficient response and power for the analyses, the order in which the scales appeared was not randomized. The final section of the surveys contained questions concerning respondent characteristics, such as age, gender, and well-being at work.

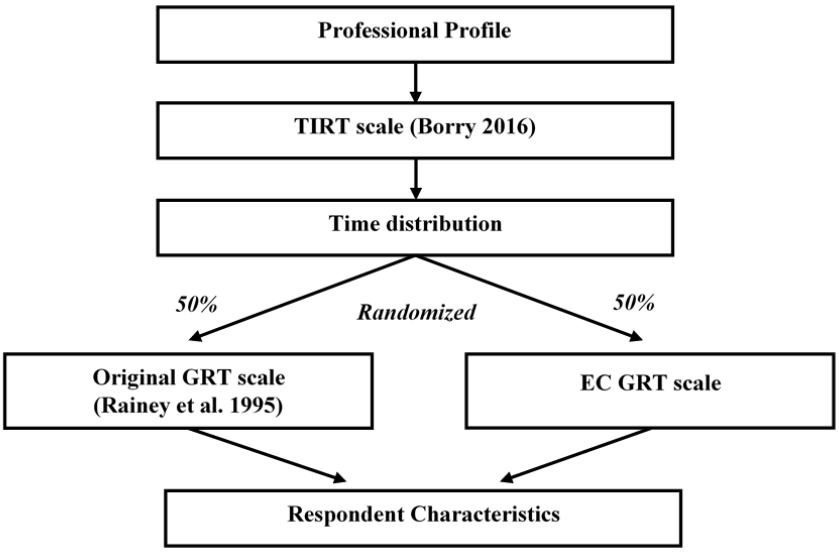

Figure 1: Survey flow T_0

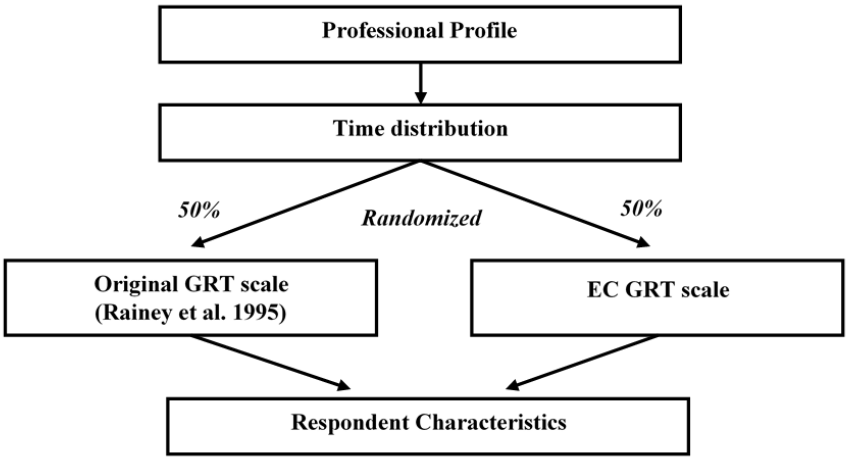

Figure 2: Survey flow at T_1

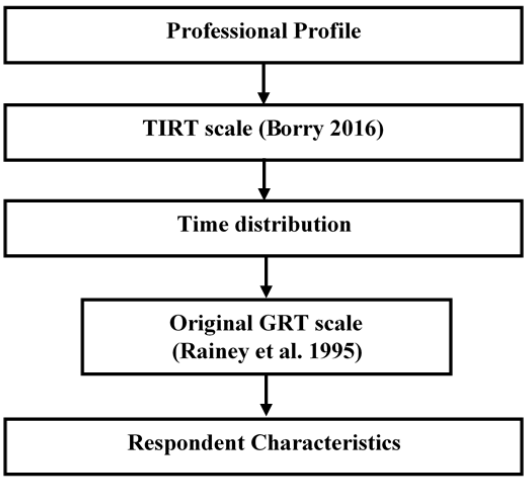

Figure 3: Survey flow at T_2

Survey Development and Distribution

In order to translate the TIRT scale and both versions of the GRT scale, this article used the combined translation technique proposed by Cha et al. (2007). To collect the data, electronic surveys were sent out to the various stakeholder groups (i.e. physicians, head nurses, nurses, paramedical staff, and secretaries). Representatives of each stakeholder group were involved in the development of the survey instrument during workshop sessions organized for each group separately. Before distributing the survey among physicians and hospital staff, the final survey instrument was evaluated and approved by two independent ethical committees. The first survey was completed by 383 respondents (13% of hospital staff), the second by 836 respondents (29%), and the third by 157 respondents (5%). The low response rate of the final measurement can be explained by the limited time interval between the second and the final survey. Due to these relatively low response rates, a sample representability check was performed by comparing the composition of the samples to the composition of the population. As shown in Appendix 1, the samples are a reliable representation of the hospital population. To interpret survey results, multiple workshops and

interviews with stakeholder representatives were performed in the weeks after survey completion. The entire study was conducted between 2018 and 2019.

TIRT scale

As the TIRT scale is a relatively new instrument in public administration research, and especially in a healthcare context, this article first evaluated the TIRT scale using confirmatory factor analyses (CFA) to assess construct validity. Cronbach's alpha was employed to check internal consistency of the TIRT construct.

GRT and TIRT validity test

The primary aim of this article is to investigate whether the GRT and TIRT scales capture a change in red tape levels due to a major change in administrative processes. Therefore, we will test whether red tape levels did in fact increase as a consequence of this change with a set of self-reported control variables in the survey, such as the proportion of time spent on administration relative to communication with colleagues and patient interaction, well-being at work variables, amount of hours worked on a weekly basis, and comparative red tape perception questions on subtasks. The well-being at work variables include job satisfaction, motivation, job qualification, workload, and risk of burnout. These variables were included since multiple studies have shown the significant relationship between well-being at work variables and red tape perceptions (Feeney, 2012; Feeney and Bozeman, 2009; Giauque et al., 2012; Steijn and van der Voet, 2019). All well-being at work variables were measured using a 7-point Likert scale ranging from very low to very high.

To check on a lower level whether red tape levels changed, we employ a specific set of core administrative work tasks and asked respondents for each task to indicate whether they perceived more or less red tape compared to the situation before the new EHR. The identified administrative tasks include reporting, registrations, orders (for labs, scans, etc.),

administration concerning medication (e.g. ordering, admission registration), and anamnesis. Furthermore, we added tasks on which the new EHR should not have an impact as a control mechanism to check respondents' choice patterns (i.e. briefings and quality control). The comparative red tape questions concluded by asking respondents whether they experienced in general more or less red tape in their job activities compared to the situation before the new EHR. Finally, extensive workshops with hospital staff and an extensive series of observations were performed at various, carefully selected departments in the hospital both before and after the implementation of the EHR.

Second, we will examine whether the GRT scale and/or the TIRT scale pick up the presumed change in red tape levels. To statistically test whether GRT and TIRT results significantly change over time we used t-tests.

GRT versus EC Conceptualization of Red Tape

As previously mentioned, we hypothesize the perceptual difference between the GRT and EC definitions influences red tape evaluations. The operationalization of both versions is shown in Appendix 2. To test whether there is a definition effect, leading to variations in ratings on the 0-10 scale, each respondent was randomly assigned to one of the two definitions. First, a t-test was performed to test the equality of means of GRT and EC levels. Second, regression models were estimated using the ordinary least squares (OLS) estimation technique while controlling for profession, age, gender, campus, experience, and the timing of the survey (i.e. before or after the EHR implementation). To make sure no structural difference between the GRT and EC respondents exists, leading to different perceptions on red tape levels not attributable to the displayed definition, we employed the percentage of time spent on administration (identical to all respondents) as a check for balance between the two groups.

Results

The descriptive results of the three surveys, presented in Table 1, show that on average respondents perceive high levels of red tape in the hospital when using the GRT scale. With averages above 7, the GRT levels in this setting are among the highest observed in red tape research (Borry, 2016; Kaufmann and Feeney, 2012; Welch and Pandey, 2006). This stresses the importance of the red tape problem in hospitals. Remarkably, 30% to 35% of work-related time was reported to be spent on administration, while only 41% to 46% of time is spent on patient care and communicating to patients' family.

TIRT scale

To check internal consistency of the TIRT scale as a construct variable, we first calculated Cronbach's alpha (0.67 at T_0 and 0.75 at T_2). With a value at T_0 below and at T_2 slightly above the threshold value of 0.70, the internal consistency of the TIRT construct is relatively poor. The reason for this finding can be found in the correlation coefficients of the three indicators. The burden indicator is weakly correlated with the unnecessary (0.19 at T_0 and 0.38 at T_2) and ineffective (0.30 at T_0 and 0.47 at T_2) indicators, while the unnecessary and ineffective indicators are correlated more strongly (0.73 at T_0 and 0.67 at T_2). Since internal consistency of the TIRT scale in the first survey is too low and model convergence was not achieved, the CFA, presented in Figure 4, only employs the data of T_2.

Table 1: Descriptive statistics

		T_0					T_1					T_2				
		N	Mean	Std. Dev.	Min	Max	N	Mean	Std. Dev.	Min	Max	N	Mean	Std. Dev.	Min	Max
GRT		383	7.2167	1.7966	0	10	836	7.1734	1.7855	0	10	157	7.0828	1.8535	0	10
	GRT definition	191	7.1257	1.8764	0	10	418	7.1938	1.8068	2	10	157	7.0828	1.8535	0	10
	EC definition	192	7.3073	1.7136	1	10	418	7.1531	1.7657	0	10	/	/	/	/	/
PercPatients		316	46.6118	19.8981	0	94.2857	796	41.3960	21.4414	0	100	133	42.8217	18.1029	0	87.5000
PercComm		316	23.7750	12.8905	1.3333	73.5294	796	24.0480	13.7983	0	83.3333	133	22.6318	12.6219	0	64.5161
PercAdmin		316	29.6132	15.1288	2.8571	86.9565	796	34.5561	17.2788	0	100	133	34.5465	16.8250	0	93.5484
TIRT		383	3.6049	1.8680	0	10	/	/	/	/	/	157	4.1274	1.9834	0	10
	Burdensome	383	4.8695	2.6186	0	10	/	/	/	/	/	157	5.2357	2.6120	0	10
	Unnecessary	383	2.6162	2.2305	0	10	/	/	/	/	/	157	3.1147	2.2100	0	10
	Ineffective	383	3.3290	2.3516	0	10	/	/	/	/	/	157	4.0318	2.4532	0	10

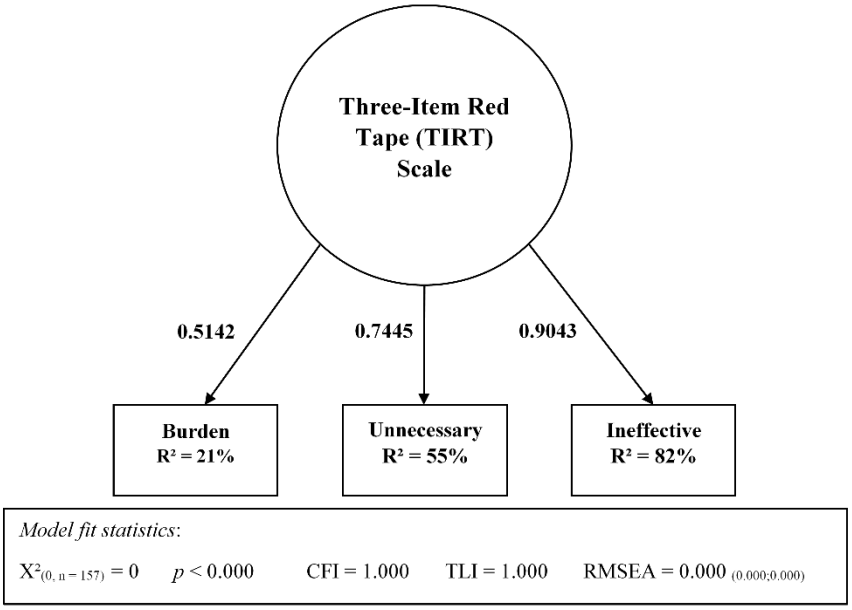

Figure 4: CFA of TIRT scale

The results at T_0 suggest that the three characteristics of red tape are not suited to be treated as one construct variable in the context of this research. This is in line with recent findings of Kaufmann et al. (2020) and Hattke et al. (2020). While the results at T_2 show an acceptable Cronbach's alpha, the variance explained in the burden indicator is low (21%). When constructing two variables instead of one construct variable, Cronbach's alpha of the unnecessary – ineffective construct increases to 0.84 at T_0 and 0.80 at T_2. As already mentioned, it is likely in this context that certain administrative tasks (e.g. allergies registration) are considered as burdensome, while they are necessary in safeguarding patient safety and quality of care. Consequently, these results seem to indicate that at least two separate constructs are required to measure red tape in a valid way.

GRT and TIRT analyses over time

To examine the second hypothesis in this article, we will examine the evolution in the GRT and TIRT scales after which we will demonstrate red tape levels did increase in the organization as a consequence of the new EHR using multiple control variables. Table 2 shows the impact of the implementation on the red tape variables in the first panel, and the

impact on the control variables in the second panel. The results indicate that the GRT averages do not vary as much over time as expected. The differences between the averages before and after the implementation are therefore not significant ($p<0.1$). Since we employed two different red tape definitions accompanying the scale (i.e. GRT and EC definition), the analysis was repeated for each version separately. This however leads to the same conclusion.

When looking at the TIRT scale a different conclusion can be made. As shown in Table 2, the red tape average measured by the TIRT scale does increase significantly. When repeating the same analysis for each of the TIRT indicators, it can be observed that both the ineffective and unnecessary indicators increase significantly but the increase in the burdensome indicator is not statistically significant. From the workshops after survey completion, it could be derived that hospital staff had to complete more administrative tasks than before the implementation. However, using the fragmented software, hospital staff endured a higher mental burden due to the free text entry as they had to think of all relevant observations themselves. The EHR on the other hand was less mentally demanding since it required them to check all possible observations one by one. While the administrative tasks became more burdensome in time, they got less mentally burdensome to complete which partly offsets the time increase in the burdensome indicator. In other words, while procedures and rules are perceived as more unnecessary and definitely more ineffective, mainly due to the fact that several check boxes are not applicable to the patient and the addition of new features to the EHR not deemed necessary to safeguard quality of care, the overall burden of these tasks did not increase significantly.

Table 2: Before-after analysis

	T_0	T_1	*p* value T_0 vs T_1	T_2	*p* value T_0 vs T_2
Red Tape variables					
GRT scale	7.22	7.17	.6590	7.08	.4143
GRT definition	7.13	7.19	.6901	7.08	.8133
EC definition	7.31	7.15	.2885	/	/
TIRT scale	3.60	/	/	4.13	.0039
Burdensome	4.87	/	/	5.24	.1403
Unnecessary	2.62	/	/	3.11	.0184
Ineffective	3.33	/	/	4.03	.0019
Control variables					
% of time spent on administration	29.61	34.56	.0000	34.55	.0024
Job satisfaction	5.07	4.59	.0000	4.63	.0001
Motivation	5.60	5.26	.0000	5.14	.0000
Job qualification	5.53	5.36	.0049	5.27	.0032
Risk of burnout	3.76	4.02	.0061	4.07	.0315
Workload	5.43	5.47	.5762	5.36	.4665
Hours worked(weekly)	37.81	38.23	.5103	36.02	.0451

The contrasting results of both scales raises questions on which scale gives the most reliable representation of the impact of the EHR on red tape in the hospital. Therefore, several control variables were added in the analyses to be able to conclude on this dilemma. As shown in panel 2 of Table 2, the percentage of work-related time spent on administrative tasks increases significantly from 30% to 35% in both follow-up measurements. Furthermore, each of the well-being at work variables deteriorates significantly over time. An exception in the well-being at work variables is workload, which remains at the same level as well as the hours worked in a week, which even decrease slightly in the final measurement. This would explain why the burdensome indicator of the TIRT scale did not increase significantly over time since hospital staff does not experience a higher overall burden in their job, but have the feeling they have to perform less necessary and effective tasks.

To supplement the aforementioned control variables, comparative red tape questions were used to be able to judge whether red tape levels did in fact increase or not. As shown in

Table 3, respondents indicated experiencing significantly more red tape in the key administrative work tasks. To check whether respondents did not indicate increased red tape in all tasks out of frustration with the new EHR, we added two tasks that should not be impacted by the new EHR (i.e. briefings and quality control). As presented in Table 3, the averages of both tasks are around 4, indicating red tape levels remained the same and the above answers can be deemed valid. Finally, an overall comparative red tape question, which can be interpreted as a proxy for the evolution in red tape, also indicates red tape did in fact increase in the organization as almost 75% of respondents indicated an increase of red tape. The observations, interviews, and workshops with hospital staff confirmed these results.

Table 3: Comparative red tape questions

	T_2 compared to T_0
Administrative tasks	
Reporting	5.16
Registrations	5.52
Orders (e.g. labs, scans, etc.)	5.01
Medication registration	5.13
Anamnesis	5.21
Tasks on which the EHR should not have an impact	
Briefings	4.06
Quality control	4.38
Overall comparative red tape question	
Overall, do you perceive more or less red tape compared to the situation before the implementation of the EHR?	5.49

Note: 7-point Likert scale from 'a lot less' to 'a lot more'

As demonstrated by the results of the control variables and the comparative red tape questions, it is clear that the implementation of the new EHR did not lead to more formalization, but instead lead to an increase in organizational red tape. This sudden, substantial increase in red tape levels offers the opportunity to test the validity of both the GRT and the TIRT scales by examining whether this increase is reflected in the results of each scale. From the analyses of the GRT and TIRT scales, it is clear that only the TIRT scale increases

significantly over time and should thus be considered as the more valid instrument of the two.

The most plausible explanation for this finding is that the GRT scale does not measure red tape accurately as it is too difficult for staff members to evaluate red tape in the entire organization. In this case, the GRT scale does not detect the change in organizational red tape due to the bottom-up approach compared to the frequently used top-down approach. Managers could be in a better position to judge red tape in the entire organization in contrast to staff members. This would also explain the large differences in GRT and TIRT averages within each measurement, which have not been observed in the past, as the TIRT scale focuses on work environment red tape and the GRT scale on organizational red tape (Borry, 2016; Kaufmann et al., 2020, 2021). The GRT scale could thus be an instrument that is ultimately suited to conduct interorganizational research using a top-down approach, but not appropriate for intraorganizational research following the bottom-up approach.

GRT versus EC conceptualization

To examine whether respondents perceive different levels of red tape depending on the definition used in the GRT scale, each respondent was assigned randomly to one of two definitions. The unbiased percentages of time spent on administration, which preceded the randomization in the survey, were used to perform a check for balance with t-tests between the two groups. As Appendix 3 demonstrates, the group that was allocated to the EC definition is highly comparable to the group allocated to the GRT definition.

To investigate whether a different definition leads to different results on the single indicator scale, a t-test was performed in each of the two measurements. The results of the t-tests clearly indicate no significant difference in average ratings between the GRT and EC definitions ($p<0.1$). This analysis was repeated for each stakeholder group separately in Appendix 4, which leads to an identical conclusion. Additionally, we examined the data further by estimating two regression models using the ordinary least squares (OLS) estimation

technique to control for professional and demographic characteristics, and the timing of the survey. The first model controls for profession, gender, age, campus, and the timing of the survey (before or after EHR implementation). The second model additionally controls for experience. The estimated models, shown in Table 4, clearly indicate there is no significant definition effect, even at the significance level of 0.1. This would suggest that although the two definitions differ significantly in wording and emphasis, respondents do not alter their red tape perceptions based on the definition provided to them. Post-survey workshops indicated respondents already have an overall idea of what red tape is and to what extent it affects their clinical practice, and that because of this general belief the definition provided to them did not influence red tape perceptions. The results of the regression analyses also provide additional support for the previously discussed finding that there is no significant difference over time in GRT averages when controlling for respondents' professional and demographic characteristics and the definition provided to them.

Table 4: Regression models using OLS

VARIABLES	(1) GRT	(2) GRT
Profession		
Nurse	.0159	-.0447
	(.1558)	(.1594)
Paramedical staff	-.6767***	-.7453***
	(.2277)	(.2289)
Physician	-.3919*	-.3701*
	(.2083)	(.2097)
Secretary	-.3797	-.4670*
	(.2558)	(.2581)
Others	-.4818*	-.4683*
	(.2627)	(.2625)
Gender		
female	-.1651	-0.1678
	(.1427)	(.1435)
Age		
26 to 35	.2154	.1389*
	(.1649)	(.1836)
36 to 45	.0168	-.1893

	(.1721)	(.2019)
46 to 55	.01407	-0.2870
	(.1743)	(.2185)
Above 55	-.0388	-0.4707*
	(.2136)	(.2730)
Campus		
B	-.1809	-.1568
	(.1404)	(.1408)
C	.2363*	.2101
	(.1387)	(.1380)
D	.2214	.2558
	(.1695)	(.1706)
E	.3176***	.3110***
	(.1132)	(.1120)
Experience		
1 to 5 years		.0285
		(.2288)
6 to 10 years		.1835
		(.2449)
11 to 20 years		.4179*
		(.2434)
21 to 30 years		.4039
		(.2664)
Above 30 years		.7016
		(.3080)
Timing		
After implementation(T_1)	-.2910	-.0646
	(.1132)	(.1144)
Definition EC	**.0193**	**.0171**
	(.1017)	**(.1015)**
Constant	7.2454***	7.2303***
	(.2655)	(.3044)
Observations	1219	1219
R-squared	.0325	.0414

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

While not being the scope of this study, the results of the regression analyses also indicate significant differences in the GRT ratings between the different stakeholders in the hospital. Paramedical staff, physicians, and 'others' experience significantly less red tape compared to head nurses. When controlled for experience as well, this is also the case for secretaries. Furthermore, the results indicate that perceived red tape decreases with age, but increases with

experience. This can be explained by the relatively high retraining rates towards the healthcare sector in the Belgian labor market. In our samples at T_0 and T_1, for example, 23% and 31% of the respondents older than 55 have less than 10 years of experience.

Conclusion

This article examines both the GRT shortcomings and the potential of the TIRT scale as an alternative using a survey experiment in a Belgian general hospital. First, we tested the validity of both scales using a before-after analysis in a repeated cross-sectional research design by examining whether each scale is able to capture a major increase in red tape. Analyses showed that the change in red tape is only reflected in the TIRT scale, while the average of the GRT scale remained constant. The GRT scale, in this context, thus fails to act as an instrument to measure the impact of organizational change or policy measures on organizational red tape. A possible reason for this finding could be the intraorganizational bottom-up approach of this study in contrast to the more frequently used interorganizational top-down approach.

Second, we examined whether the definition used to describe red tape leads to variations in perceived levels of red tape. By randomly assigning either the original GRT definition or the European Commission's conceptualization of red tape to each respondent, we are able to conclude that although the two definitions differ in wording and emphasis, respondents reflect on the same underlying red tape concept when evaluating red tape in their organization.

Overall, we can conclude the GRT scale is not sensitive to changes in organizational red tape and the definition accompanying the scale does not influence results. The results seem to advocate for a multiple indicator scale, preferably one in which the term 'red tape' is not mentioned since respondents tend to make a cognitive shortcut to their own belief of what red tape means. Further longitudinal and experimental research is needed to examine the validity of the GRT scale and its alternatives in a multitude of research contexts. Furthermore, we discovered the internal consistency of the TIRT construct is relatively poor. Therefore, in

future red tape research in a healthcare context, the Job-Centered Measure of Van Loon et al. (2016) should be explored to examine red tape perceptions and as an instrument to measure the impact of organizational change on red tape.

Supplementary material:

Appendix 1: Survey samples versus population

		T_0	T_1	T_2	Population
Profession					
	Head nurses	10%	7%	10%	3%
	Nurses	58%	55%	62%	68%
	Paramedical staff	13%	8%	9%	7%
	Physician	4%	16%	7%	8%
	Secretaries	7%	5%	4%	6%
	Others	6%	8%	8%	8%
Gender					
	female	82%	75%	78%	87%
	male	18%	25%	22%	13%
Age					
	Under or equal to 25	10%	7%	7%	6%
	Between 26 and 35	29%	25%	25%	30%
	Between 36 and 45	27%	28%	25%	26%
	Between 46 and 55	23%	26%	22%	23%
	Above 55	12%	14%	20%	15%
Campus					
	A	22%	26%	27%	22%
	B	19%	17%	13%	11%
	C	18%	17%	19%	13%
	D	9%	13%	11%	9%
	E	50%	48%	50%	44%
Experience					
	Under 1 year	7%	7%	7%	7%
	Between 1 and 5 years	24%	20%	18%	16%
	Between 6 and 10 years	22%	19%	14%	21%
	Between 11 and 20 years	26%	25%	27%	24%
	Between 21 and 30 years	14%	18%	18%	18%
	Above 30 years	7%	12%	16%	15%
Respondents		383	836	157	2 863

Appendix 2: Operationalization of GRT scale with two definitions

Original General Red Tape (GRT) scale as presented in the survey (50% random sample)

If red tape is defined as "burdensome rules and procedures that have negative effects on the organization's effectiveness", how would you assess the level of red tape in your organization?

Almost no red tape	0	1	2	3	4	5	6	7	8	9	10	A great deal of red tape
	0	0	0	0	0	0	0	0	0	0	0	

General Red Tape (GRT) scale with European Commission's definition as presented in the survey (50% random sample)

If red tape is defined as "the time and costs incurred in meeting legal obligations to provide information on your activities, either to public or private authorities, which are done solely because of this legal obligation", how would you assess the level of red tape in your organization?

Almost no red tape	0	1	2	3	4	5	6	7	8	9	10	A great deal of red tape
	0	0	0	0	0	0	0	0	0	0	0	

Appendix 3: Balance check

		Percentage of time spent on administration					
		T_0			T_1		
		EC	GRT	*p* value	EC	GRT	*p* value
Profession							
	Head nurse	/	/	/	42.61	40.40	.5598
	Nurse	28.12	28.94	.6519	36.80	35.10	.2527
	Paramedical staff	31.06	30.33	.8784	27.14	32.01	.2821
	Physician	23.90	22.74	.8842	24.67	26.44	.5117
	Secretary	100	100	1.000	100	100	1.000
	Others	28.09	28.61	.7726	36.44	35.10	.3666
Gender							
	female	30.06	28.90	.5350	35.53	35.48	.9742
	male	28.12	33.61	.1912	31.13	33.16	.4225
Age							
	Under or equal to 25	19.94	27.51	.1332	40.45	33.26	.0769
	Between 26 and 35	30.12	30.60	.8671	33.50	34.36	.7064
	Between 36 and 45	27.59	31.33	.2847	33.66	35.42	.4688
	Between 46 and 55	35.02	26.02	.0324	35.03	35.83	.7485
	Above 55	30.67	34.13	.5134	31.95	34.26	.5496
Campus							
	A	33.09	30.73	.5487	33.46	31.59	.4525
	B	24.23	28.96	.2419	34.47	38.29	.2340
	C	29.18	27.32	.6004	32.97	33.71	.8154
	D	31.60	31.78	.9779	32.66	37.66	.1778
	E	31.64	32.54	.7064	34.88	37.00	.2539
Experience							
	Under 1 year	30.48	32.62	.8094	35.53	32.99	.5210
	Between 1 and 5 years	32.78	29.54	.4251	35.54	37.22	.5317
	Between 6 and 10 years	28.00	32.01	.2813	34.09	36.61	.4029
	Between 11 and 20 years	29.08	27.61	.6142	32.99	32.11	.7030
	Between 21 and 30 years	26.82	27.46	.8759	32.18	33.29	.7091
	Above 30 years	34.58	29.31	.4161	38.17	37.06	.7818
Overall		29.66	29.56	.9546	34.25	34.93	.5781
n		192	191		400*	396*	

**Note: due to 40 respondents who completed the GRT scale but not the 'percentage of time' question, the number of respondents differs from Table 1. Conducting the analyses with this sample leads to the same conclusions.*

Appendix 4: GRT averages with EC or GRT definition

		Red tape averages					
		T_0			T_1		
		EC	GRT	*p* value EC vs GRT	EC	GRT	*p* value EC vs GRT
Profession							
	Head nurse	7.48	7.47	.9765	7.04	7.45	.2592
	Nurse	7.36	7.42	.7944	7.27	7.40	.4305
	Paramedical staff	7.17	6.56	.2613	6.45	6.72	.5756
	Physician	6.75	5.50	.3723	7.16	7.00	.6114
	Secretary	7.70	6.53	.1138	7.18	6.59	.2860
	Others	7.07	6.22	.4520	7.11	6.83	.5490
Gender							
	female	7.25	7.06	.3350	7.11	7.22	.4197
	male	7.58	7.47	.8086	7.28	7.12	.5588
Age							
	Under or equal to 25	6.75	6.84	.8308	7.16	7.41	.5163
	Between 26 and 35	7.60	7.53	.8428	7.11	7.32	.3579
	Between 36 and 45	7.40	7.16	.4922	7.06	7.17	.6702
	Between 46 and 55	7.00	7.09	.8349	7.18	7.15	.8767
	Above 55	7.27	6.74	.3249	7.31	7.00	.4377
Campus							
	A	7.11	6.83	.5136	7.20	7.21	.9537
	B	6.86	6.86	.9922	7.11	6.93	.5536
	C	7.61	7.64	.9258	7.07	7.10	.9108
	D	7.37	6.52	.2220	7.10	7.66	.1000
	E	7.30	7.41	.7058	7.35	7.25	.5783
Experience							
	Under 1 year	6.41	6.56	.8660	6.68*	7.85*	.0046
	Between 1 and 5 years	7.30	7.02	.3962	6.94	7.01	.7858
	Between 6 and 10 years	7.25	7.33	.8577	6.93	7.25	.2298
	Between 11 and 20 years	7.70	7.18	.1215	7.37	7.06	.2296
	Between 21 and 30 years	7.21	7.00	.6808	7.21	7.24	.9142
	Above 30 years	7.20	7.22	.9782	7.60	7.25	.3689
Overall		7.31	7.13	.3232	7.15	7.19	.7421
n		192	191		418	418	

References

Arendsen R, Peters O, ter Hedde M and van Dijk J (2014). Does e-government reduce the administrative burden of businesses? An assessment of business-to-government systems usage in the Netherlands. *Government Information Quarterly* 31(1): 160-169.

Borry E (2016). A New Measure of Red Tape: Introducing the Three-Item Red Tape (TIRT) Scale. *International Public Management Journal* 19(4): 573-593.

Bozeman B (1993). A Theory of Government "Red Tape". *Journal of Public Administration Research and Theory: J-PART* 3(3): 273-303.

Bozeman B (2000). *Bureaucracy and red tape.* Englewood Cliffs: Prentice Hall.

Bozeman B (2012). Multidimensional Red Tape: A Theory Coda. *International Public Management Journal* 15(3): 245-265.

Bozeman B and Feeney M (2011). *Rules and red tape: a prism for public administration theory and research.* Armonk: Sharpe.

Cha E, Kim K and Erlen J (2007). Translation of scales in cross-cultural research: Issues and techniques. *Journal of Advanced Nursing* 58: 386-395.

De Jong G and Van Witteloostuijn A (2015). Regulatory Red Tape and Private Firm Performance. *Public Administration* 93(1): 34-51.

European Commission (2017). Better Regulation Toolbox. In (pp. 540). Brussels: European Commission.

Feeney M (2012). Organizational red tape: a measurement experiment. *Journal of Public Administration Research and Theory* 22(3): 427.

Feeney M and Bozeman B (2009). Stakeholder Red Tape: Comparing Perceptions of Public Managers and Their Private Consultants. *Public Administration Review* 69(4): 710-726.

Giauque D, Ritz A, Varone F and Anderfuhren-Biget S (2012). Resigned but Satisfied: The Negative Effect of Public Service Motivation and Red Tape on Work Satisfaction. *Public Administration* 90(1): 175-193.

Hattke F, Hensel D and Kalucza J (2020). Emotional Responses to Bureaucratic Red Tape. *Public Administration Review* 80(1): 53-63.

Henderson A and Borry E (2020). The emotional burdens of public service: rules, trust, and emotional labour in emergency medical services. *Public Money and Management*.

Jacobsen C and Jakobsen M (2018). Perceived Organizational Red Tape and Organizational Performance in Public Services. *Public Administration Review* 78(1): 24-36.

Kaufmann W and Feeney M (2012). Objective Formalization, Perceived Formalization and Perceived Red Tape. *Public Management Review* 14(8): 1195-1214.

Kaufmann W, Ingrams A and Jacobs D (2020). Rationale and process transparency do not reduce perceived red tape: evidence from a survey experiment. *International Review of Administrative Sciences*.

Kaufmann W, Ingrams A and Jacobs D (2021). Being Consistent Matters: Experimental Evidence on the Effect of Rule Consistency on Citizen Red Tape. *The American Review of Public Administration* 51(1): 28-39.

Kaufmann W, Taggart G and Bozeman B (2019). Administrative Delay, Red Tape, and Organizational Performance. *Public Performance and Management Review* 42(3): 529-553.

Kaufmann W and Tummers L (2017). The negative effect of red tape on procedural satisfaction. *Public Management Review* 19(9): 1311-1327.

Luton L (2007). Deconstructing Public Administration Empiricism. *Administration and Society* 39(4): 527-544.

Migchelbrink K and Van de Walle S (2020). Increasing the cost of participation: red tape and public officials' attitudes toward public participation. *International Review of Administrative Sciences.*

OECD (2007). Cutting Red Tape: Comparing Administrative Burdens across Countries. In (pp. 180). Paris: OECD Publications.

Oh Y and Lee K (2020). External control mechanisms and red tape: testing the roles of external audit and evaluation on red tape in quasi-governmental organizations. *International Review of Administrative Sciences.*

Pandey S and Marlowe J (2014). Assessing Survey-Based Measurement of Personnel Red Tape With Anchoring Vignettes. *Review of Public Personnel Administration* 35(3): 215-237.

Poel K, Marneffe W, Bielen S, Van Aarle B and Vereeck L (2014). Administrative simplification and economic growth: A cross country empirical study. *Journal of Business Administration Research* 3(1): 45.

Rainey H, Pandey S and Bozeman B (1995). Research note: Public and private managers' perceptions of red tape. *Public Administration Review* 55(6): 567.

Steijn B and van der Voet J (2019). Relational job characteristics and job satisfaction of public sector employees: When prosocial motivation and red tape collide. *Public Administration* 97(1): 64-80.

Van Loon N (2017). From Red Tape to Which Performance Results? Exploring the Relationship Between Red Tape and Various Dimensions of Performance in Healthcare Work Units: Red Tape and Performance. *Public Administration* 95(1): 60-77.

Van Loon N, Leisink P, Knies E and Brewer G (2016). Red Tape: Developing and Validating a New Job-Centered Measure. *Public Administration Review* 76(4): 662-673.

Walker R and Brewer G (2009). Can Management Strategy Minimize the Impact of Red Tape on Organizational Performance? *Administration and Society* 41(4): 423-448.

Welch E and Pandey S (2006). E-Government and Bureaucracy: Toward a Better Understanding of Intranet Implementation and Its Effect on Red Tape. *Journal of Public Administration Research and Theory* 17(3): 379-404.