

**Perceptual Bias and Public Programs:
The Case of the United States and Hospital Care**

Forthcoming, *Public Administration Review*

Kenneth J. Meier
Distinguished Scholar in Residence
American University,
Cardiff University,
and
Leiden University
kmeier@american.edu

Austin P. Johnson
PhD Candidate
Texas A&M University
john31860@tamu.edu

and

Seung-Ho An
Assistant Professor
University of Arizona
seunghoan@arizona.edu

Suggested citation: Meier, Kenneth J., Austin P. Johnson, and Seung-Ho An. 2019. "Perceptual Bias and Public Programs: The Case of the United States and Hospital Care." *Public Administration Review*. 79(6): 820-828.

Replication data and Stata do-file available at Harvard Dataverse:
<https://doi.org/10.7910/DVN/NRQPZX>.

Abstract

This study examines whether the public holds biased perceptions of public organizations (hospitals) in the United States and whether organizations get credit for positive results from program evaluations. Using an experimental design that replicates Hvidman and Andersen's (2016) Danish study, we find no negative public sector biases in the US, but organizations are not given any credit for positive program evaluations. These results hold in two experimental replications. We discuss the implications of our findings for the measurement of public perceptions of government programs and for effective democratic governance.

Practitioner Points

- Performance appraisal results generated by the agency do not appear to affect public evaluations of the agency.
- US public hospitals are not perceived more negatively than private hospitals in terms of efficiency, effectiveness, red tape or benevolence.
- Experiments on citizen evaluation need to be replicated in specific agency and national contexts.
- Practitioners need to be especially careful about arguments related to efficiency and whether reliable measures of efficiency are possible.

The ability of citizens to evaluate the performance of government and public programs is a prerequisite of democratic governance (Achen and Bartels 2017; Pande 2011). Citizens are expected to judge government policies and programs and vote accordingly (Downs 1957), thus, linking the recent performance management movement (Moynihan 2008; Sanderson 2001) to larger questions of governance. Although a modest literature has documented the match-up or lack of match-up between citizen assessments of programs and official government measures of performance (see Andrews, Boyne and Walker 2006; Song and Meier 2018 and the references cited therein), that literature cannot distinguish between whether a lack of congruence results because citizens have different objectives than government programs or citizens are unable to judge the quality of government programs. Of particular concern is the inability to evaluate programs or the unwillingness to do so that might result from among other causes such as the basic perceptual biases that psychologists have found that most individuals have (Tversky and Khaneman 1981). Similar biases in public perceptions have been found in the growing literature on behavioral public administration (Baekgaard and Serritzlew 2016; Battaglio, Belardinelli, Belle, and Cantarelli 2018; Belle, Cantarelli and Belardinelli 2018; Charbonneau and Van Ryzin 2015; James et al. 2016; James 2011a, 2011b; Marvel 2015a; 2015b; Olsen 2017; Riccucci, Van Ryzin, and Lavena 2014; Riccucci, Van Ryzin, and Li 2016).

A recent article in *Public Administration Review* by Hvidman and Andersen (2016) found negative views of the public sector per se. That experiment had students in Denmark to evaluate hospitals on several dimensions and showed that students had a more negative view of the public sector in terms of efficiency, effectiveness, and red tape, and a modestly more positive

assessments on benevolence.¹ Hvidman and Andersen frame the Danish context as enhancing the theoretical and substantive significance of the findings suggesting that Denmark is one of the least likely nations to harbor anti-public sector biases. Health care in Denmark is a government-funded entitlement and corresponds with an elaborate public sector social net for its citizens; the Danish government spends 54.8% of GNP, the third highest among OECD countries (only France and Finland spend more; OECD 2018). Bureaucrat bashing in Denmark is relatively modest according to Hvidman and Andersen, so they present Denmark as a hard case. If negative perceptions of the public sector exist in Denmark, the logic suggests that they are likely to exist in other countries with larger private sectors and greater anti-government rhetoric. They specifically state “given the negative framing of the public sector in the United States (Bok 2001), effects might be even more pronounced” (Hvidman and Andersen 2016: 117). The implication is that perceptions in the US might be so biased as to render citizen assessments of government performance of questionable validity (see also Marvel 2015a).

In contrast to the welfare state orientation of Denmark, the United States has a relatively small government sector (37.7% of GDP, OECD 2018, the ninth smallest in the OECD) and thus a general preference for private sector delivery of services. Bureaucrat bashing has long been a tradition in the United States dating back to at least the Jacksonian contention that government jobs were simple enough that anyone could do them (White 1954; for more contemporary examples see Goodsell 2015, 2018; Marvel 2015a; Rölle 2017). The recent “drain the swamp” rhetoric is only the latest example. Related to the substantive area of health care and hospitals, the US spends far more per capita on health care (\$10,348, Centers for Medicare and Medicaid

¹ In Hvidman and Anderson (2016), effectiveness and benevolence are not statistically significant at the traditional statistical significance level.

Services [CMS] 2016) than other countries; and nearly \$2 trillion of the \$3.3 trillion annual expenditures come from federal, state and local governments (CMS 2016; “Total US Government Spending” 2018). Approximately one-third of US health expenditures are for hospital care (National Center for Health Statistics 2016). To the extent that the US public holds sector biases in regard to health care, as a result, could greatly affect a major area of public policy.

In addition to these substantive policy reasons for investigating the Hvidman and Andersen (2016) results in the US, there are important theoretical reasons for reexamining question in the US. Public administration as an institution adapts to national context, and this means that determining how generalizable the findings of public sector bias are is important. This is especially valuable given the major differences between the US and Denmark in terms of health policy (universal care versus limited care), health expenditures (\$10,348 versus \$4,553 per capita), and the sizes of the public sector (37.6% of GDP versus 54.8%). Among developed countries, the comparison of Denmark and the US essentially presents a natural “least similar systems” design (King, Keohane and Verba 1994), an optimal design to determine if findings can be generalized.

Public Administration as a Design Science

Public administration is a science of the artificial; it is concerned with not just what is, but what might be (Simon 1969). Although knowledge for its own sake has value, public administration is generally concerned with applications, with the transfer of scholarly knowledge to the practical world of the public administrator. Such an orientation establishes a higher

standard for the reliability and validity of research. After all, if knowledge is sought merely for its own sake, then whether that knowledge is actually true remains only an intellectual concern. If knowledge is sought for the purpose of changing policy, programs, systems or people, then it is incumbent on scholars of public administration to be certain of their findings and thus their recommendations to the world of practice.

Of crucial importance in the transfer of scholarly knowledge to the world of practice is establishing the boundary conditions (Walker et al. 2019). Does X (the new program, etc.) work all the time or does it only work when certain other conditions are met (levels of funding, specified procedures, in areas of high social capital, etc.)? It may be that the success of X is probabilistic, that is, it works some of the time, but that this probability is not random but rather a function of context either internal to X (that is how the program is designed) or external to X (that is, in certain types of environments). Contextual factors clearly matter. We know that the willingness of citizens to pay taxes varies a great deal from one country to another (Casey and Scholz 1991; Scholz and Pinney 1995). Program take-up rates for welfare programs are significantly lower in the US than in Western Europe (Nickell 1997; Remler, Rachlin, and Glied 2001). Determining the boundary conditions of a program or policy (*or in perceptions of sector bias*) and assessing how general a finding is in any policy area requires replication, the repeating of scientific studies to determine the veracity of a claim.

Replicating experimental studies in different national contexts are one way to bridge what has been termed the micro-macro gap in public administration research (Moynihan 2018). Experiments focus on internal validity and establishing causal claims and thus are ideal for examining questions of individual attitudes and behavior. They sacrifice external validity and thus often do not address the big questions that concern public administration, the macro

dimension (Moynihan 2018). One way to bridge this gap is the replication of experimental work in different contexts, in the present case, different national contexts.

Although replicating studies in different national contexts has a long history in psychology (see Rokeach 1960 with the seminal work of Hofstede 2001) spawning an entire field of cultural psychology (Kitayama and Cohen 2007; applied to organizations see Smith et al. 2002), public administration scholarship is just beginning such work. First, there is some evidence that generic macro factors such as unemployment (Olsen 2017), the state of the economy (Powell and Whitten 1993; Singer 2011), organizational past performance (Olsen 2017), the performance of others (Charbonneau and Van Ryzin 2015), institutional logics (Ngoye, Sierre, and Ysa 2018), institutional arrangements (Baldwin 2018), and social capital (Beugelsdijk and Van Schaik 2005) influence individual attitudes and behaviors. Second, such national contexts as types of democracy (Anderson and Guillory 1997), quality of public administration (Ariely 2014), and government age and change (Armigeon and Guthmann 2013) are associated with differences in public satisfaction with government. Individual taxpayer compliance, one of the basic needs of all governing systems, is also associated with levels of corruption (DeBacker, Heim and Tran 2015), provision of public goods (Hallsworth et al. 2017), and social norms (Luttmer and Singhal 2014). Olsen et al. (2019) demonstrate that behaviors linked to corruption correlate with macro estimates of country level corruption, and Kim et al. (2013) show distinct cross-national patterns of public service motivation. Third, there is suggestive evidence that evaluations of health care programs vary across nations (Blendon et al. 1990), regions (Jha et al. 2008) and sectors (Peters et al. 2007) that are associated with different national contexts. All three sets of literature indicate that investigating the area of public sector bias in different national contexts can provide new insights for public administration.

The Original Experiment

To illustrate the importance of experimental replication and to address the important question of citizens' abilities to evaluate programs, we replicated the published Danish experiment by Hvidman and Andersen (2016). That survey experiment investigated whether the public perceived that public organizations are inefficient and burdened by administrative procedures (and also whether public organizations are perceived as more benevolent or more effective). A total of 148 subjects (political science undergraduate students) were randomly exposed to different vignettes about a hospital in a survey. The experimental condition was whether or not the hospital was identified as a public hospital or a private hospital; a second experimental condition was whether or not the hospital created an external performance appraisal system to assess its performance (with the performance appraisal reporting positive results). The objective was to determine if a performance appraisal process improved the respondents' evaluations of public programs. The experiment found that respondents rated the public hospital more negatively than the private hospital in terms of efficiency and red tape; the differences in the respondents' ratings in regard to effectiveness and benevolence were not statistically significant. Further, the use of performance management did not significantly change the respondents' assessment of the public hospital (or the private hospital).

Given the relationship between bureaucracy and democratic governance, the Hvidman-Andersen experiment poses two troubling findings. First, the addition of positive performance information had no influence on how individuals evaluated program performance (in either the public or the private sector). This finding fundamentally challenges the notion of a citizenship-

electoral base that objectively evaluates government performance as the first step in a political accountability process. Second, the perceptions are biased against public organizations indicating even if public organizations are superior in program performance that citizens will not recognize that fact.

We replicated the Hvidman and Andersen experiment in the United States with two additional differences (what Walker et al. 2019:7 refer to as an “empirical generalization” replication) in addition to the different country context. Rather than using students, we initially implemented the experiment with an on-line sample of 385 adults using Amazon MTurk (a second replication used students, see below). The first difference, that is, using adults proved to make no difference since similar results were found in the US student sample. For the second difference, an assessment of the statistical power needed in the experimental design indicated the need for a larger sample size (Walker et al. 2019; 9-10); that difference ended up being relatively unimportant given the results of both of the US experiments.²

As Hvidman and Andersen (2016) suggest, one might expect the perceptual biases to be stronger in the US context than in Denmark. First, Denmark operates an elaborate welfare state using a corporatist model of government, business, and labor cooperation. This governance process blurs the difference between government and the private sector framing them as partners with labor in the process (Schmitter 1974) which should reduce perceptual bias. In contrast the size of the public sector is and has been a major political issue in the US since the realignment of

²The replication was embedded within a larger experiment with 1,039 subjects. In addition to the replication experiment, we also included an experimental group using the cue of nonprofit hospitals (given the size of the nonprofit hospital sector in the United States) and whether the use of performance information was proactive or reactive (that is, whether the hospital did it on its own initiative). The respondents from those portions of the experiment were excluded from this replication assessment. The nonprofit results were not statistically different from either the public sector or the private sector results.

1896 with major partisan differences (Key 1955, Sundquist 2010). Contentions about the superiority of the private sector abound (Savas 1982) despite mixed evidence (Hodge 2000). This perception is reinforced by the frequent “bureaucrat bashing” engaged in by various politicians where government is blamed for all the country’s shortcomings and by the US enthusiasm for the New Public Management (Pollitt and Bouckaert 2004). One of the underlying premises of the NPM is that government can be improved by adopting the methods of business including the greater use of the private sector to deliver government services (Hood 1991). As Marvel (2015a: 143) notes, “from birth to death, the idea that public sector organizations are inefficient, wasteful, and inferior to private sector organizations is consistently drilled into Americans’ heads.” In the specific area of health care, access to health care is a highly contentious and partisan issue in the United States as demonstrated by the politics of the Affordable Care Act (Gitterman and Scott 2011; Haeder and Weimer 2013; Jacobs and Metler 2018). In Denmark universal access to health care has been guaranteed by law since 1973 when Denmark abolished state subsidized insurance systems and established a single payer system using government funds. As one commentary on health care in Denmark (Vallgård, Krasnik, Vrangbæk 2001: 10) concluded: “Danish welfare politics in general, and health care politics in particular, are characterized by consensus regarding basic institutional structures. In the years since the Second World War, political parties on all sides have continued to support the idea that access to health care should be independent of ability to pay or place of residence.”

Second, using adults with greater experience with hospitals and the health care sector rather than undergraduate students should also increase the difference in perceptions. Perceptions of bias need to come from somewhere, either personal experiences or exposure to information or arguments of others. On such dimensions, one might expect there to be

differences between college students and an older sample of the population who have more experience with hospitals (but see the replication results below).

In addition to these designed differences, the context of hospital care in the US and Denmark differ and might also influence the findings. As noted, access to health care in Denmark is universal, and the majority of health care in Denmark is provided by public hospitals. Health care in the US is on a fee-for-service basis, and costs are generally covered by third party payers (insurance companies, government). Public hospitals are a minority of the hospitals in the US. At the same time while government involvement in health care is more extensive in Denmark than the US, in both countries government plays a major role in funding and regulating the quality of health care. We should expect that using a larger sample, in a country where public-private differences are emphasized and highly salient, with an adult population would increase public sector bias and, therefore, that the biases would be larger in the US than in Denmark.

Any discussion of a replication that uses identical vignettes as this one does should follow three steps (see Walker et al. 2019; Walker, James, and Brewer 2017 on replications). First, the analyst should determine whether the randomization itself was successful, the “balance test” across experimental groups. Second, the concepts measured in the replication need to be as reliable and valid as those in the initial experiment. Whether concepts measure the exact same thing across national contexts is a question that needs to be established empirically rather than assumed (Jilke et al. 2017). If a concept has a different meaning in the new context, then any statistical comparisons become questionable. As a key example, the meaning of “public service motivation” appears to vary across nations with distinctly different meanings in authoritarian versus democratic regimes (see Kim et al. 2013). Third, a parallel testing of the impact of the

experimental conditions needs to be presented.

Table 1 assesses the balance of respondents across the four groups: public experimental (that is, performance appraisal), private experimental, public control, and private control. Five variables used for this assessment are age, gender, race, religious service attendance, and ideology (the Danish experiment used age, gender, ideology, preference for the public sector, and public service motivation). Both experiments found balance, that is, the f-test of the difference of means across the groups did not have any statistically significant differences at the 0.05 level of significance; such a balance test is standard for determining whether the randomization was successful in generating groups that were not different from each other.

[Table 1 About Here]

The Danish experiment used nine items to measure performance, and a factor analysis confirmed that the nine items generated four distinct factors that the authors termed effectiveness, red tape, efficiency, and benevolence. Their initial hypotheses were that the public hospital would rate lower on effectiveness and efficiency and higher on red tape and benevolence. Table 2 presents the factor loadings as well as the Cronbach's alphas for the US respondents (Table 2a) and the Danish respondents (Table 2b).

[Table 2 About Here]

The differences in the factor analyses merit discussion. First, the effectiveness factor appears to be more reliable in the US than in Denmark (Cronbach's alpha of 0.90 versus 0.67). The differences for the Danish and US effectiveness item loadings are such that they are statistically distinct, that is, statistical tests indicate they are unlikely to be from the same sample. Basic rules of thumb for reliability (see Kline 2000, 13; DeVellis 2012, 109-110) suggest that alphas above 0.9 are excellent, between 0.8 and 0.9 are good, between 0.7 and 0.8 are acceptable,

between 0.6 and 0.7 are questionable, between 0.5 and 0.6 are poor and less than 0.5 are unacceptable. Second, in contrast, the efficiency factor is far more reliable in Denmark than in the US (Cronbach's alpha 0.63 versus 0.29); the loadings are also statistically different from each other. Even though a Cronbach's alpha of 0.63 would be termed "questionable," the Danish alpha is a marked improvement over the US alpha of 0.29. Efficiency as measured by these two items in the US cannot be considered a reliable concept, and any findings in regard to it need to be treated with skepticism.³ Third, both red tape and benevolence are measured consistently in the two countries and have similar levels of reliability. The respective Danish and US alphas are 0.80 and 0.76 for red tape and 0.86 and 0.93 for benevolence. This comparison of measurement reliability indicates that we should be fairly confident that assessments of red tape and benevolence are comparing the same concept in the two countries; the comparison on effectiveness should also be reasonable, but the efficiency comparisons could well be limited by the differences in measurement reliability.

The initial experimental findings in Hvidman and Andersen (2016) showed the direct effects of both being a public hospital and also the performance management cue. Table 3 compares the US results of this experiment (top of table) with the Danish results (bottom of table). Hvidman and Andersen find that public organizations are rated significantly lower than private hospitals on both red tape and efficiency. Although public organizations rank lower on effectiveness and higher on benevolence than private hospitals, those impacts were not statistically significant. The performance information cue had no impact on the ratings in

³The alpha coefficient for efficiency in the student sample was 0.47. Together these findings raise an interesting question about perceptual measures of efficiency in the United States in general. The term might be so value laden in the United States that it lacks empirical referents such as lower costs or balanced budgets. This is a measurement issue that merits further investigation.

Denmark; that is, hospitals' ratings do not change if they seek a performance appraisal that then produces positive findings.

[Table 3 About Here]

The US findings are *not* consistent with the Danish findings. Public hospitals are rated no differently than private hospitals on any of the performance indicators. While signs of the coefficients are similar to those for the Danish sample, none of the estimates are statistically different from zero. In short, the hypothesized public sector bias that was modestly supported (two of four cases) in Denmark is not present in the US sample despite the much larger sample size in the US. The effect sizes in the US even if they were statistically significant are, in fact, miniscule. Reexamining the Danish results, however, indicates that the two significant findings for efficiency and red tape⁴ have t-scores of approximately 2.0, that is, they are statistically significant at the 0.05 level but not much more than that.

Although the impact of the performance management cue in the US is similar to the Danish findings, statistically insignificant in three of the cases, the one striking exception is that US hospitals that seek an outside evaluation of their performance are perceived as more benevolent than those that do not. Because the questions ask if the hospital is interested in the “well-being of patients” and “acts in the interest of patients,” the effort to acquire an outside evaluation of the hospital and the resulting positive performance information may be associated with increased perceptions of benevolence.

[Table 4 About Here]

The final analysis in Hvidman and Andersen (2016, 117) interacted the use of

⁴Both studies reverse code red tape so that higher scores indicate less red tape and all four indicators are then signed consistently with each other.

performance appraisal with the public sector variable to determine if performance appraisal mattered more in the public sector than the private sector. None of those interaction coefficients in Table 4 were statistically significant leading them to conclude that performance management or the use of an outside consultant to provide [positive] performance information was ineffective in both types of hospitals. The US results (in the top half of table 4) are consistent with the Danish conclusion. None of the relationships for red tape, efficiency, effectiveness and benevolence are statistically significant at the 0.05 level.

Given the differences in our findings and the original experiment, a valid question is whether this resulted from using a sample of adults rather than college students? To determine if this was the case, we replicated the experiment again with 416 undergraduates at a large public university. The results of this experiment (see appendix tables A1-A4) were consistent with those in the original US sample. Students did not have significant negative stereotypes of public hospitals, and the addition of performance information had no real influence on the results. The reliability of efficiency as a concept remained problematic. The differences between the current study and Hvidman and Andersen cannot be attributed to the use of students versus adults.

Before discussing the implications of this replication effort, a recapitulation of the similarity and differences in the experimental results is merited. First, the introduction of performance appraisal had no impact on performance assessments in either Denmark or the United States. Similarly, performance appraisal did not matter in either the public or the private sector. Second, the concept of efficiency as measured in Denmark did not generate a reliable measure in the United States (although it was relatively unreliable in Denmark also). Cross-national efficiency comparisons, a monumental task in general, will be limited by the lack of a reliable and valid measure of efficiency in the United States. Third, there was no evidence of

public sector biases in the United States unlike the Danish case where the public sector was perceived as less efficient and more burdened by red tape.

Discussions and Conclusions

The motivation for this study was Hvidman and Andersen's (2016) findings in Denmark that the public viewed public organizations as more inefficient and more burdened by red tape than their private counterparts. The addition of positive performance information which in theory should increase perceptions of effectiveness and perhaps efficiency had no influence on these perceptions. Hvidman and Andersen suggest that such findings, particularly the bias against public sector organizations are likely to be even stronger in the United States with its greater history of bureaucrat bashing and general preference for private sector delivery of services.

This study replicated the Hvidman-Andersen experiment twice – once with a sample of adults via Mturk and once with a sample of university students. Both replications in the US did not find that individuals rated public sector organizations any differently than private sector organizations on measures of effectiveness, efficiency, benevolence, and red tape. In terms of the use of performance information, both US experiments confirmed the Danish results. Positive performance information did not affect how individuals rated either the public or the private hospitals in the study.

Focusing first on the differences in the two studies, the public sector bias in Denmark versus the lack of bias in the US, the theoretical expectation was that the biases would be greater in the US than in Denmark. The null findings in the US were, thus, unexpected and merit

discussion. First, the difference in the findings cannot be attributed to an inadequate sample size; power tests suggested a larger sample was needed and was implemented in the US. Second, the differences cannot be attributed to the use of students in Denmark given that the second replication in the US also used a corresponding student population (see Walker et al. 2018 on diversity in experimental pools).

What factors might account for the differences in the US and Danish results? In the control groups individuals are given no information about the performance of the hospital being evaluated. The original objective was that this would permit respondents to let their evaluations be colored by their general perceptions of public and private organizations. It could be the case that US respondents do not apply such biases, if they have them, to hospitals. Although the performance of hospitals is a highly complex and multidimensional concern, a superficial comparison using national data from the Center for Medicare and Medicaid Services in the US (CMS 2018) shows virtually identical average performance by sector. Using the most recent data available for 2009-2011, the average 30-day hospital readmission rates for the three tracked incidents (heart disease, heart failure and pneumonia) were 21.12% for public hospitals and 21.10% for private ones. The average mortality rates for the same diagnoses (2008-2011) were 13.26% for public hospitals and 13.12% for private. Finally, the overall five-star ratings (where 5 is best) for 2017 were 2.96 stars for public hospitals and 2.83 for private. Although these are gross comparisons and do not consider the range of functions that hospitals perform, there is no clear evidence that either set of hospitals is superior to the other (see also Cheon et al. 2019). One explanation for the initial null findings, therefore, is that the public perceives no differences simply because on average there are no differences and thus the priming had no influence.

A second explanation might be that in a mixed delivery system, stereotypes become less

relevant because it is ambiguous as to what sector is involved. The literature defines a public organization in at least three ways – ownership, source of funding, and degree of government control (Rainey and Bozeman 2000). Both public and private hospitals are subject to extensive government control both in terms of cost regulations and in terms of quality; the existence of the extensive Hospital Compare data base is only one example. Similarly, both public and private hospitals have a substantial portion of their revenues from government sources either via Medicare or Medicaid or from government grants. Only a few very rare hospitals that are highly specialized and do not accept insurance payments can avoid these financial entanglements. That leaves only the ownership distinction which in this case might be far less relevant than the other aspects of publicness. Bozeman's (1987) contention that all organizations are public could well be applied to US hospitals.

Our study offers a case where the public perception of public and private organizations does not differ, which adds an episode to the theoretical story on public-private differences and the evaluation of government services. We acknowledge that there are many empirical and logical studies that have found and argued for differences in public and private organizations (Boyne, Poole, and Jenkins 1999; Bullock, Stritch, and Rainey 2015; Rainey and Bozeman 2000; Rainey, Pandey and Bozeman 1995; Wamsley and Zald 1973). The experiments conducted here suggest that careful probing of these general attitudes is needed to determine when and under what conditions public programs are viewed as inferior to private ones.

In addition to different national contexts, replication studies are needed in different types of public services. Both the Danish experiments by Hvidman and Andersen (2016) and our US experiments use (public and private) hospitals as a case for analysis, and both studies found that the performance appraisal does not affect citizens' evaluations on organizational performance,

regardless of being public or private. As noted, it might well be the case that US experiences with hospitals do not generate different patient treatments and, thus, reinforce biases. Research on other types of services such as education (Meier et al. 2015), water and waste services (Bel, Fageda, and Warner 2010), recreational services (Eagles 2009; Talmage, Anderson, and Searle 2018), long-term care (Amirkhanyan et al. 2019), and so forth are needed. Agencies providing various types of services tend to have different organizational characteristics; and, therefore, the way public service professionals and citizens interact with each other – properties of bureaucratic encounters – can exhibit different features. These can lead to different public perceptions about agencies working in the particular service areas.

The findings on perceptual measurements of performance appraisal are perhaps more troublesome and merit discussion. Democratic theory expects that the public will reward policy and political actions that benefit citizens and penalize those that harm them. In both the Danish case and the US case, hospitals were given no credit for engaging in evaluations and reporting positive results. While this finding is consistent with an extensive literature on negativity bias which finds that people discount good news and focus on bad news, that cannot be the entire story. If only bad news gains the attention of the public, then all public programs (and probably all private ones) would have negative evaluations. In fact, many public programs have positive reputations, and an extensive literature in education shows that the public has some skill at discerning good schools from bad ones (see Charbonneau and Van Ryzin 2012).

Future research needs to address the lack of impact of performance-related information. One possibility is that the current assessment provided only general information about improvement; future research might include performance information that permits more quantitative or comparative information. Another possibility is that respondents discounted the

performance information because it originated from the hospital seeking a consultant and thus, the source of the information might not have been credible. Again, future research could examine various sources of performance information from official government statistics to subjective assessments by citizens.

Equally important for future research is the need for better measurement. Our factor analysis results suggest that reliability of measures for organizational efficiency and effectiveness are inconsistent across the two countries, although benevolence and red tape measures are measured with reasonable consistency. The efficiency measure in the US failed to reach even minimum levels of reliability whether the sample was the MTurk adults or the US college students. This indicates that public perceptions in a variety of contexts can significantly differ, which may affect the reliability of measurements. Measuring the performance of public organizations can be inherently subjective and challenging due to goal ambiguity. Ongoing efforts to replicate and examine the reliability of performance measures in various contexts, therefore, would be a valuable step in developing better measures and improving the quality of such measures.

Of particular concern in terms of measurement is the lack of reliability of the measure of efficiency in the US context. The discussions of public versus private provision of services consistently emphasize the potential efficiencies of the private sector (see Savas 1982) although there has always been mixed evidence on the superiority of the private sector (see Hodge 2000). The problematic efficiency findings might result because there are no differences in hospital performance in the US and the perceptions match the bureaucratic encounters. Alternatively, the concept of efficiency might have little meaning to the consumer where the costs of services are generally covered by third party payers (government, insurance companies). Or of even greater

concern, the concept of efficiency may either have little meaning for consumers or the meaning that consumers attach to the term is different from that of the scholarly community. At a minimum, new and reliable measures of efficiency are needed to conduct experimental research in this area.

Finally, we would highlight the potential importance of pursuing replication. These endeavors produce evidence based on different events and episodes in various contexts, which reveal previously missing parts of the universe for assessing public services. We would encourage future scholars to conduct more replication in terms of quantitative analysis, experiments, and measurement evaluation, as well as journals to consider publishing studies with null results.

References

- Achen, Christopher H., and Larry M. Bartels. 2017. *Democracy for Realists: Why Elections Do not Produce Responsive Government*. Princeton, NJ: Princeton University Press.
- Amirkhanyan, Anna A., Stephen B. Holt, Austin M. McCrea, and Kenneth J. Meier. 2019. Managing Racial Diversity: Matching Internal Strategies with Environmental Needs. *Public Administration Review* 79 (1): 69-81.
- Anderson, Christopher J., and Christine A. Guillory. 1997. Political Institutions and Satisfaction with Democracy: A Cross-national Analysis of Consensus and Majoritarian Systems. *American Political Science Review* 91 (1): 66-81.
- Andrews, Rhys, George A. Boyne, and Richard M. Walker. 2006. Subjective and Objective Measures of Organizational Performance: An Empirical Exploration. Pp. 14-34 in George A. Boyne, et al. *Public Service Performance: Perspectives on Measurement and Management*. London: Cambridge University Press.
- Ariely, Gal. 2013. Public Administration and Citizen Satisfaction with Democracy: Cross-national Evidence. *International Review of Administrative Sciences* 79 (4): 747-766.
- Armingeon, Klaus, and Kai Guthmann. 2014. Democracy in Crisis? The Declining Support for National Democracy in European Countries, 2007–2011. *European Journal of Political Research* 53 (3): 423-442.
- Baekgaard, Martin, and Søren Serritzlew. 2016. Interpreting Performance Information: Motivated Reasoning or Unbiased Comprehension. *Public Administration Review* 76 (1): 73-82.
- Baldwin, Elizabeth. 2018. Exploring How Institutional Arrangements Shape Stakeholder Influence on Policy Decisions: A Comparative Analysis in the Energy Sector. *Public Administration Review* <https://doi.org/10.1111/puar.12953>
- Battaglio Jr, R. Paul, Paolo Belardinelli, Nicola Bellé, and Paola Cantarelli. 2018. Behavioral Public Administration ad fontes: A Synthesis of Research on Bounded Rationality, Cognitive Biases, and Nudging in Public Organizations. *Public Administration Review*. <https://doi.org/10.1111/puar.12994>
- Bel, Germà, Xavier Fageda, and Mildred E. Warner. 2010. Is Private Production of Public Services Cheaper than Public Production? A Meta-Regression Analysis of Solid Waste and Water Services. *Journal of Policy Analysis and Management* 29 (3): 553-577.
- Bellé, Nicola, Paola Cantarelli, and Paolo Belardinelli. Prospect Theory Goes Public: Experimental Evidence on Cognitive Biases in Public Policy and Management Decisions. *Public Administration Review* 78, no. 6 (2018): 828-840.

- Beugelsdijk, Sjoerd, and Ton Van Schaik. 2005. Social Capital and Growth in European Regions: An Empirical Test. *European Journal of Political Economy*, 21(2): 301-324.
- Blendon, Robert J., Robert Leitman, Ian Morrison, and Karen Donelan. 1990. Satisfaction with Health Systems in Ten Nations. *Health Affairs* 9 (2): 185-192.
- Bok, Derek. 2001. *The Trouble with Government*. Cambridge MA: Harvard University Press.
- Boyne, George, Michael Poole, and Glenville Jenkins. 1999. Human Resource Management in the Public and Private Sectors: An Empirical Comparison. *Public Administration* 77(2): 407-420.
- Bozeman, Barry. 1987. *All Organizations Are Public: Bridging Public and Private Organizational Theories*. San Francisco: Jossey-Bass Inc.
- Bullock, Justin B., Justin M. Stritch, and Hal G. Rainey. 2015. International Comparison of Public and Private Employees' Work Motives, Attitudes, and Perceived Rewards. *Public Administration Review* 75(3): 479-489.
- Casey, Jeff T., and John T. Scholz. 1991. Beyond Deterrence: Behavioral Decision Theory and Tax Compliance. *Law and Society Review*, 25 (4): 821-843.
- Centers for Medicare and Medicaid Services. 2016. National Health Expenditure Data. <https://www.cms.gov/Research-Statistics-Data-and-Systems/Statistics-Trends-and-Reports/NationalHealthExpendData/NationalHealthAccountsHistorical.html>, accessed August 27, 2018.
- Centers for Medicare and Medicaid Services. 2018. Hospital Compare Overall Hospital Rating. <https://www.medicare.gov/hospitalcompare/About/Hospital-overall-ratings.html>, access October 14, 2018.
- Charbonneau, Étienne, and Gregg G. Van Ryzin. 2012. Performance Measures and Parental Satisfaction with New York City Schools. *American Review of Public Administration* 42 (1): 54-65.
- Charbonneau, Étienne, and Gregg G. Van Ryzin. 2015. Benchmarks and Citizen Judgments of Local Government Performance: Findings from a Survey Experiment. *Public Management Review* 17 (2): 288-304.
- Cheon, Ohbet, Miyeon Song, Austin M. McCrea and Kenneth J. Meier. 2019. Health Care in America: The Relationship between Subjective and Objective Assessments of Hospitals. *International Public Management Journal*. doi/full/10.1080/10967494.2018.1553808
- DeBacker, Jason, Bradley T. Heim, and Anh Tran. 2015. Importing Corruption Culture from Overseas: Evidence from Corporate Tax Evasion in the United States." *Journal of Financial Economics*, 117(1): 122-138.

- DeVellis, Robert F. 2012. *Scale Development: Theory and Applications*. Los Angeles: Sage.
- Downs, Anthony. 1957. *An Economic Theory of Democracy*. New York: Harper and Row.
- Eagles, Paul FJ. 2009. Governance of Recreation and Tourism Partnerships in Parks and Protected Areas." *Journal of Sustainable Tourism* 17 (2): 231-248.
- Gitterman, Daniel P., and John C. Scott. 2011. "Obama Lies, Grandma Dies": The Uncertain Politics of Medicare and the Patient Protection and Affordable Care Act. *Journal of Health Politics, Policy and Law* 36 (3): 555-563.
- Goodsell, Charles T. 2015. *The New Case for Bureaucracy: A Public Administration Polemic*. Thousand Oaks, CA: Sage.
- Goodsell, Charles T. 2018. "Bureaucrat Bashing." PP. 128-131 in Jay M. Shafritz, ed. *Defining Public Administration*. New York: Routledge.
- Haeder, Simon F., and David L. Weimer. 2013. You Can't Make Me Do It: State Implementation of Insurance Exchanges under the Affordable Care Act." *Public Administration Review* 73 (s1): S34-S47.
- Hallsworth, Michael, John A. List, Robert D. Metcalfe, and Ivo Vlaev. 2017. The Behavioralist as Tax Collector: Using Natural Field Experiments to Enhance Tax Compliance. *Journal of Public Economics*, 148(1): 14-31.
- Hodge, Graeme. 2000. *Privatization: An International Review of Performance*. New York: Routledge.
- Hofstede, Geert H. 2001. *Culture's Consequences: Comparing Values, Behaviors, Institutions and Organizations across Nations*. Thousand Oaks, CA: Sage Publications.
- Hood, Christopher. 1991. A Public Management for all Seasons?" *Public administration*, 69(1): 3-19.
- Hvidman, Ulrik and Simon Calmar Andersen. 2016. Perceptions of Public and Private Performance: Evidence from a Survey Experiment. *Public Administration Review* 76 (1): 111-120.
- Jacobs, Lawrence R., and Suzanne Mettler. 2018. When and How New Policy Creates New Politics: Examining the Feedback Effects of the Affordable Care Act on Public Opinion. *Perspectives on Politics* 16 (2): 345-363.
- James, Oliver. 2011a. Performance Measures and Democracy: Information Effects on Citizens in Field and Laboratory Experiments. *Journal of Public Administration Research and Theory* 21(3):

399-418.

James, Oliver. 2011b. Managing Citizens' Expectations of Public Service Performance: Evidence from Observation and Experimentation in Local Government. *Public Administration* 89(4): 1419-1435.

James, Oliver, Sebastian Jilke, Carolyn Petersen, and Steven Van de Walle. 2016. Citizens' Blame of Politicians for Public Service Failure: Experimental Evidence about Blame Reduction through Delegation and Contracting. *Public Administration Review* 76(1): 83-93.

Jha, Ashish K., E. John Orav, Jie Zheng, and Arnold M. Epstein. 2008. Patients' Perception of Hospital Care in the United States. *New England Journal of Medicine* 359 (18): 1921-1931.

Jilke, Sebastian, Nicolai Petrovsky, Bart Meuleman, and Oliver James. 2017. Measurement Equivalence in Replications of Experiments: When and Why it Matters and Guidance on How to Determine Equivalence. *Public Management Review* 19 (9): 1293-1310.

Luttmer, Erzo FP, and Monica Singhal. 2014. Tax Morale. *The Journal of Economic Perspectives*, 28(4): 149-168

Key Jr, Valdimer O. 1955. A Theory of Critical Elections. *The Journal of Politics* 17 (1): 3-18.

Kim, Sangmook., Wouter Vandenaabeele, Bradley E. Wright, Lotte Bøgh Andersen, Francesco Paolo Cerase, Robert K. Christensen, Céline Desmarais, M. Koumenta, Peter Leisink, B. Liu, J. Palidauskaitė. 2013. Investigating the Structure and Meaning of Public Service Motivation across Populations: Developing an International Instrument and Addressing Issues of Measurement Invariance. *Journal of Public Administration Research and Theory*, 23(1): 79-102.

King, Gary, Robert O. Keohane, and Sidney Verba. *Designing Social Inquiry: Scientific Inference in Qualitative Research*. Princeton university press, 1994.

Kitayama, Shinobu and Dov Cohen. 2011. *Handbook of Cultural Psychology*. New York: The Guilford Press.

Kline, Paul. 2000. *The Handbook of Psychological Testing* (2nd ed.). London: Routledge.

Marvel, John D. 2015a. Public Opinion and Public Sector Performance: Are Individuals' Beliefs about Performance Evidence-based or the Product of Anti-public Sector Bias? *International Public Management Journal* 18 (2): 209-227.

Marvel, John D. 2015b. Unconscious Bias in Citizens' Evaluations of Public Sector Performance. *Journal of Public Administration Research and Theory* 26 (1): 143-158.

Meier, Kenneth, Simon Calmar Andersen, Laurence J. O'Toole Jr, Nathan Favero, and Søren C. Winter. 2015. Taking Managerial Context Seriously: Public Management and Performance in

- US and Denmark Schools. *International Public Management Journal* 18(1): 130-150.
- Moynihan, Donald P. 2008. *The Dynamics of Performance Management*. Georgetown University Press.
- Moynihan, Donald P. 2018. A Great Schism Approaching? Towards a Micro and Macro Public Administration. *Journal of Behavioral Public Administration* 1 (1), 1-8. DOI 10.30636/jbpa.11.15
- National Center for Health Statistics. 2016. Centers for Disease Control and Prevention. <https://www.cdc.gov/nchs/fastats/health-expenditures.htm>, accessed August 27, 2018.
- Nickell, Stephen. 1997. Unemployment and Labor Market Rigidities: Europe versus North America. *The Journal of Economic Perspectives* 11(3): 55-74.
- Ngoye, Benard, Vicenta Sierra, and Tamyko Ysa. 2018. Different Shades of Gray: A Priming Experimental Study on How Institutional Logics Influence Organizational Actor Judgment. *Public Administration Review*. doi/full/10.1111/puar.13006
- OECD (Organization for Economic Cooperation and Development. 2018. *General Government Spending*. <https://data.oecd.org/gga/general-government-spending.htm>, accessed 7/14/2018.
- Olsen, Asmus Leth. 2017. Compared to What? Experimental Evidence on Social and Historical Reference Points in Performance Evaluation. *Journal of Public Administration Research and Theory* 27(4): 562-580.
- Olsen, Asmus Leth, Frederik Hjorth, Nikolaj Harmon, and Sebastian Barfort. 2019. Behavioral Dishonesty in the Public Sector. *Journal of Public Administration Research and Theory* <https://doi.org/10.1093/jopart/muy058>.
- Pande, Rohini. 2011. Can Informed Voters Enforce Better Governance? Experiments in Low-Income Democracies. *Annual Review of Economics* 3 (1): 215-237.
- Peters, Ellen, Nathan Dieckmann, Anna Dixon, Judith H. Hibbard, and C. K. Mertz. 2007. Less Is More in Presenting Quality Information to Consumers. *Medical Care Research and Review* 64 (2): 169-190.
- Pollitt, Christopher, and Geert Bouckaert. 2004. *Public Management Reform: A Comparative Analysis*. New York: Oxford University Press.
- Powell Jr, G. Bingham and Guy D. Whitten. 1993. A Cross-national Analysis of Economic Voting: Taking Account of the Political Context.” *American Journal of Political Science*, 37 (May): 391-414.
- Rainey, Hal G., and Barry Bozeman. 2000. Comparing Public and Private Organizations:

Empirical Research and the Power of The A Priori. *Journal of Public Administration Research and Theory* 10(2): 447-470.

Rainey, Hal G., Sanjay Pandey, and Barry Bozeman. 1995. Research Note: Public and Private Managers' Perceptions of Red Tape. *Public Administration Review* 567-574.

Remler, Dahlia K., Jason E. Rachlin, and Sherry A. Glied. 2001. What Can the Take-Up of Other Programs Teach Us about How to Improve Take-Up of Health Insurance Programs? No. w8185. *National Bureau of Economic Research*

Riccucci, Norma M., Gregg G. Van Ryzin and Cecilia F. Lavena. 2014. Representative Bureaucracy in Policing: Does It Increase Perceived Legitimacy? *Journal of Public Administration Research and Theory* 24(3): 537-551.

Riccucci, Norma M., Gregg G. Van Ryzin and Huafang Li. 2016. Representative Bureaucracy and the Willingness to Coproduce: An Experimental Study. *Public Administration Review* 76(1): 121-130.

Rölle, Daniel. 2017. Mass Media and Bureaucracy-bashing: Does the Media Influence Public Attitudes towards Public Administration? *Public Policy and Administration* 32 (3): 232-258.

Rokeach, Milton. 1960. *The Open and the Closed Mind*. New York: Basic Books.

Sanderson, Ian. 2001. Performance Management, Evaluation and Learning in 'Modern' Local Government. *Public Administration* 79 (2): 297-313.

Savas, Emanuel S. 1982. *Privatizing the Public Sector: How to Shrink Government*. Chatham NJ: Chatham House Publishers.

Schmitter, Philippe C. 1974. Still the Century of Corporatism? *Review of Politics* 36 (1): 85-131.

Scholz, John T., and Neil Pinney. 1995. Duty, Fear, and Tax Compliance: The Heuristic Basis of Citizenship Behavior. *American Journal of Political Science* 39(2): 490-512.

Simon, Herbert A. 1969. *The Sciences of the Artificial*. Cambridge, MA: MIT Press.

Singer, Matthew M. 2011. Who Says 'It's the economy'? Cross-national and Cross-individual Variation in the Salience of Economic Performance. *Comparative Political Studies* 44(3), 284-312.

Smith, Peter B., Mark F. Peterson, Shalom H. Schwartz and others. 2002. Cultural Values, Sources of Guidance, and their Relevance to Managerial Behavior: A 47-nation Study. *Journal of Cross-cultural Psychology* 33 (2): 188-208.

Song, Miyeon, and Kenneth J. Meier. 2018. Citizen Satisfaction and the Kaleidoscope of

Government Performance: How Multiple Stakeholders See Government Performance. *Journal of Public Administration Research and Theory* 28 (4), 489-505.

Sundquist, James. 2010. *Politics and Policy: The Eisenhower, Kennedy, and Johnson Years*. Washington, DC: Brookings Institution Press.

Talmage, Craig A., Derrick M. Anderson, and Mark S. Searle. 2018. Whither Recreation and Parks? Understanding Change in Public Institutions Through a Theory of Adaptive Publicness. *Perspectives on Public Management and Governance* 1 (2): 143-158.

Total US Government Spending. 2018. https://www.usgovernmentspending.com/federal_budget_estimate_vs_actual_2017_XXbs2n_10#usgs302, accessed August 27, 2018.

Tversky, Amos, and Daniel Kahneman. 1981. The Framing of Decisions and the Psychology of Choice. *Science* 211 (4481): 453-458.

Signild Vallgård, Signild, Allan Krasnik, and Karsten Vrangbæk. 2001. *Health Care Systems in Transition: Denmark 2001*. Copenhagen: European Observatory on Health Care Systems.

Richard M Walker, Richard M., Gene A Brewer, M. Jin Lee, Nicolai Petrovsky, and Arjen van Witteloostuijn. 2019. Best Practice Recommendations for Replicating Experiments in Public Administration. *Journal of Public Administration Research and Theory*, doi.org/10.1093/jopart/muy047

Walker, Richard M., Oliver James, and Gene A. Brewer. 2017. Replication, Experiments and Knowledge in Public Management Research. *Public Management Review* 19 (9): 1221-1234.

Walker, Richard M., M. Jin Lee, Oliver James, and Samuel MY Ho. 2018. Analyzing the Complexity of Performance Information Use: Experiments with Stakeholders to Disaggregate Dimensions of Performance, Data Sources, and Data Types. *Public Administration Review* 78 (6): 853-863.

Wamsley, Gary L., and Mayer N. Zald. 1973. The Political Economy of Public Organizations. *Public Administration Review* 33 (1): 62-73.

White, Leonard D. 1954. *The Jacksonians: A Study in Administrative History, 1829-1861*. New York: Macmillan, 1954.

Tables

Table 1: Balance across Experimental Groups

Groups		Ideology	Age	Religious Service Attendance	Gender	White
Public	Baseline (1)	2.50	37.38	1.95	0.48	0.75
	PM (2)	2.78	35.97	2.01	0.53	0.67
Private	Baseline (3)	2.46	35.15	1.88	0.53	0.77
	PM (4)	2.38	35.87	1.89	0.52	0.77
Prob. > F		0.09	0.71	0.84	0.92	0.37

Table 2: Factor Analysis of the Four Perceived Performance Dimensions

Table 2a. Survey items (the US)	Effectiveness	Red Tape	Efficiency	Benevolence
The hospital is effective.	0.79	-0.03	-0.01	0.29
The hospital is effective in accomplishing its core mission.	0.84	0.01	0.01	0.23
The hospital is effective in delivering a very good service.	0.81	-0.04	0.08	0.23
The hospital has disproportionate administrative rules and procedures.	0.01	0.70	0.00	-0.11
A high level of administrative procedures negatively affects the hospital's efficiency.	-0.06	0.72	0.04	-0.15
The hospital is effective in lowering its costs.	0.33	0.02	0.30	0.14
The hospital prioritizes its profit margin.	0.08	0.22	0.32	-0.15
The hospital is genuinely interested in the well-being of patients.	0.23	-0.11	0.00	0.85
The hospital acts in the interest of patients.	0.31	-0.08	0.00	0.85
Cronbach's alpha	0.90	0.76	0.29	0.93
Mean	0	0	0	0
Standard deviation	0.91	0.79	0.42	0.90
N	385	385	385	385
Table 2b. Survey items (Denmark)	Effectiveness	Red Tape	Efficiency	Benevolence
The hospital is effective.	0.53	-0.17	0.17	0.14
The hospital is effective in accomplishing its core mission.	0.65	-0.06	0.11	0.18
The hospital is effective in delivering a very good service.	0.51	-0.08	-0.12	0.42
The hospital has disproportionate administrative rules and procedures.	0	0.75	-0.11	-0.01
A high level of administrative procedures negatively affects the hospital's efficiency.	-0.13	0.75	-0.04	-0.02
The hospital is effective in lowering its costs.	0.09	-0.23	0.62	0.12
The hospital prioritizes its profit margin.	0.05	-0.01	0.64	-0.27
The hospital is genuinely interested in the well-being of patients.	0.15	0.04	-0.03	0.82
The hospital acts in the interest of patients.	0.08	0.01	-0.04	0.84
Cronbach's alpha	0.67	0.80	0.63	0.86
Mean	0	0	0	0
Standard deviation	0.75	0.82	0.75	0.89
N	147	147	147	147

Table 3: Perceptions of Performance: Impact of Sector and Performance Management, Regression Models (OLS)

The US	Effectiveness	Red Tape	Efficiency	Benevolence
Public Organizations (Public=1)	-0.01 (0.09)	0.06 (0.08)	-0.04 (0.04)	0.10 (0.09)
Performance Management	-0.03 (0.09)	0.04 (0.08)	0.01 (0.04)	0.26* (0.09)
Constant	0.02 (0.08)	-0.05 (0.07)	0.02 (0.03)	-0.19* (0.08)
R-Squared overall	0.00	0.00	0.00	0.02
Prob. > F	0.94	0.67	0.62	0.01
N	385	385	385	385
Denmark	Effectiveness	Red Tape	Efficiency	Benevolence
Public Organizations (Public=1)	-0.19 (0.12)	-0.28* (0.14)	-0.24* (0.12)	0.21 (0.15)
Performance Management	0.20 (0.12)	-0.01 (0.14)	0.09 (0.12)	-0.10 (0.15)
Constant	-0.01 (0.11)	0.14 (0.12)	0.07 (0.11)	-0.05 (0.13)
N	147	147	147	147

Note. * $p < 0.05$; robust standard errors in parentheses; two-tailed tests of significance; Danish study results are directly adopted from Hvidman and Anderson (2016).

Table 4: Perceptions of Performance: Moderating Impact of Performance Management, Regression Models (OLS)

The US	Effectiveness	Red Tape	Efficiency	Benevolence
Public Organizations (Public=1)	0.16 (0.12)	0.03 (0.11)	0.02 (0.06)	0.00 (0.13)
Performance Management (PM)	0.14 (0.13)	0.01 (0.11)	0.07 (0.06)	0.17 (0.13)
Public × PM	-0.33 (0.19)	0.06 (0.16)	-0.12 (0.09)	0.19 (0.18)
Constant	-0.07 (0.08)	-0.04 (0.08)	-0.01 (0.04)	-0.14 (0.10)
R-Squared overall	0.01	0.00	0.01	0.03
Prob. > F	0.36	0.82	0.51	0.01
N	385	385	385	385
Denmark	Effectiveness	Red Tape	Efficiency	Benevolence
Public × PM	-0.26 (0.24)	0.09 (0.27)	0.17 (0.25)	-0.36 (0.29)
N	147	147	147	147

Note. * $p < 0.05$; robust standard errors in parentheses; two-tailed tests of significance; Danish studies are directly adopted from Hivdman and Anderson (2016); The authors note that all constituent terms are included but not reported.

Appendix Tables (on line not for publication)

Table A1: Balance across Experimental Groups (students)

Groups		Ideology	Age	Religious Service Attendance	Gender	White
Public	Baseline (1)	3.37	19.05	1.95	0.45	0.58
	PM (2)	3.10	18.92	1.90	0.35	0.59
Private	Baseline (3)	3.38	19.18	1.98	0.41	0.72
	PM (4)	3.28	19.04	1.85	0.52	0.73
Prob. > F		0.18	0.52	0.82	0.12	0.03

Table A2. Factor Analysis of the Four Perceived Performance Dimensions.

Survey items (US students)	Effectiveness	Red Tape	Efficiency	Benevolence
The hospital is effective.	0.67	-0.17	0.02	0.25
The hospital is effective in accomplishing its core mission.	0.73	-0.06	-0.05	0.25
The hospital is effective in delivering a very good service.	0.69	-0.07	-0.04	0.26
The hospital has disproportionate administrative rules and procedures.	-0.13	0.55	0.12	-0.21
A high level of administrative procedures negatively affects the hospital's efficiency.	-0.21	0.53	0.05	-0.17
The hospital is effective in lowering its costs.	-0.06	0.08	0.44	-0.08
The hospital prioritizes its profit margin.	-0.04	0.21	0.43	-0.17
The hospital is genuinely interested in the well-being of patients.	0.28	-0.08	-0.07	0.74
The hospital acts in the interest of patients.	0.27	-0.14	-0.04	0.74
Cronbach's alpha	0.82	0.63	0.47	0.84
Eigenvalues	1.66	0.69	0.40	1.38
Mean	0	0	0	0
Standard deviation	0.82	0.65	0.54	0.80
N	416	416	416	416

Table A3: Perceptions of Performance: Impact of Sector and Performance Management, Regression Models (students)

	Effectiveness	Red Tape	Efficiency	Benevolence
Public Organizations (Public=1)	-0.03 (0.08)	-0.01 (0.06)	-0.04 (0.05)	-0.02 (0.08)
Performance Management	-0.14 (0.08)	0.04 (0.06)	-0.04 (0.05)	-0.04 (0.08)
Constant	0.08 (0.06)	-0.01 (0.05)	0.04 (0.05)	0.03 (0.07)
R-Squared overall	0.01	0.00	0.00	0.00
Prob. > F	0.20	0.85	0.62	0.87
N	416	416	416	416

Note. * $p < 0.05$; robust standard errors in parentheses; two-tailed tests of significance.

Table A4. Perceptions of Performance: Moderating Impact of Performance Management, Regression Models (students)

	Effectiveness	Red Tape	Efficiency	Benevolence
Public Organizations (Public=1)	0.03 (0.11)	0.04 (0.09)	-0.01 (0.07)	0.02 (0.10)
Performance Management	-0.08 (0.11)	0.08 (0.09)	-0.01 (0.07)	-0.01 (0.11)
Public \times Performance Management	-0.13 (0.16)	-0.09 (0.13)	-0.08 (0.11)	-0.08 (0.16)
Constant	0.05 (0.07)	-0.03 (0.06)	0.02 (0.05)	0.01 (0.08)
R-Squared overall	0.01	0.00	0.00	0.00
Prob. > F	0.29	0.85	0.70	0.90
N	416	416	416	416

Note. * $p < 0.05$; robust standard errors in parentheses; two-tailed tests of significance.