Never Change a Winning Policy?
Public Sector Performance and Politicians' Preferences for Reforms

Abstract

Despite increasing stress on performance in public sector organizations, there remains little empirical evidence on whether – and, if so, how – politicians respond to performance information. We address this research gap by linking registry statistics on school performance in Norway's 428 municipalities with data from an information experiment embedded in a survey to local politicians (in which we randomly allocate treatments manipulating information about school performance). We find that school performance bears only a weak relationship to preferences for resource-related reforms, while it strongly affects preferences for governance-related reforms, which indicates the importance of accounting for heterogeneity across alternative types of (school) reforms. Moreover, local politicians are, on average, well informed about school performance. We show that this reflects the force of local inhabitants' high information level on politicians' accountability.

Practitioner Points:
- Performance information influences reform preferences and can be crucial for politicians to support structural policy changes.

- Politicians are well informed about the relative performance of local schools, which suggests that performance statistics do not disappear into a 'black hole'.

- Indicators of school performance have little bearing on politicians' preferences for resource-related reforms, but strongly affect their position regarding governance-related reforms. This has direct implications for the design and focus of performance management strategies.

"We shall describe as a 'bureaucratic system of organization' any system of organization where the feedback process (…) does not function well, and where consequently there cannot be any quick readjustment of the programs of action in view of the errors committed."
Michel Crozier, *The Bureaucratic Phenomenon* (1964: 186-187)

As indicated by Crozier's (1964) description of bureaucratic systems, public sector organizations are often assumed to be characterized by substantial rigidity. This may leave little potential for the increasing stress on performance in public sector organizations to trigger reforms, since a rigid organization "cannot correct its behavior by learning from its errors" (Crozier 1964: 187). Moreover, organizational learning can only arise when performance information is actively sought and used by policy-makers, and the extent to which this happens remains debated (Ter Bogt 2004; Andersen and Moynihan 2016; Bjørnholt et al. 2016; Andersen and Jakobsen 2017; for a review, see Kroll 2015). From an empirical and policy perspective, this raises the important question whether – and, if so, how – policy-makers want to initiate policy reforms in response to performance information.

Most existing work on the extent and antecedents of performance information use is based on self-reports (Kroll 2015). It addresses whether and how frequently actors take into account performance information (e.g. Moynihan and Pandey 2010) or how important they think such information is to their decisions (e.g. Askim 2007). More recent scholarship also investigates how observed performance in the public sector affects preferences for, and the probability of, policy reforms. For instance, Salge (2011) shows that performance problems positively affect the search for innovations in 154 public hospitals. Nielsen (2014) similarly finds that Danish school principals prioritize issues with observable under-performance when performance is weak, while Nicholson-Crotty et al. (2017) show that performance affects the level of risk-taking by employees in US federal agencies. Finally, Askim et al. (2008: 297) highlight that performance information obtained via benchmarking "supports organizational learning and innovation" in Norwegian municipality-level policy networks. Overall, and in contrast to Crozier's (1964) pessimistic view, these studies suggest that employees in public sector organizations *do* respond to organizational results.

Still, studies linking performance to reforms should not only investigate *administrative behavior*, but also assess the preferences and/or behavior of *political decision-makers* (Heinemann et al. 2016). Currently, there is "less systematic evidence on how performance

data alter political beliefs" (Nielsen and Moynihan 2016: 3), which constitutes our first contribution. To the best of our knowledge, Nielsen and Baekgaard (2015) and George et al. (2016) are the only studies linking information on public sector performance to preferences for policy reforms using data obtained from politicians.[1] Their analysis is limited, however, to budgetary allocations and one reform option, school mergers. Instead, we allow for a broader spectrum of possible reform measures, and also investigate the accuracy of politicians' performance information. We identify the effects of performance information on politicians' reform preferences using a survey experiment whereby council members in Norway's 428 municipalities are randomly assigned to three information treatments.

Our second contribution relates to the *types of reforms* that might be triggered by public sector performance. Politicians generally have multiple reform tools at their disposal, which carry distinct levels of expected performance gains and political costs. In our education setting, especially reforms prioritizing teacher quality, school accountability and school closures are characterized by substantial political costs – relative to reforms to extend pre-school services, assistance to weak students or teacher discretion. Teacher unions tend to oppose increased emphasis on teacher quality, and instead view formal qualifications and job experience as preferred criteria for hiring and wage settlement. Stricter performance accountability likewise meets antagonism from key stakeholders including unions and public employees (Pollitt 2006; Bonesrønning 2013). Finally, proposals for school closures trigger intense media disputes and opposition from local communities during the preparation of the municipal budget.[2] Exploiting insights from recent work on negativity bias in citizens' and politicians' perception of performance (Olsen 2015a; Nielsen and Moynihan 2017), we maintain that performance information is particularly likely to affect politicians' preferences for reforms involving greater political risks (such as the risk of losing key supporters). The reason is that politically costly reforms require a stronger defense from politicians, and performance information can help provide the necessary justification (Johnsen 2012; Asatryan et al. 2017).

Our main findings can be summarized as follows. First, school performance has weak and statistically insignificant effects on politicians' preferences for *resource-related reforms*, but significantly affects preferences for *governance-related reforms*. Second, explicitly providing performance information produces at best modest changes in politicians' reform preferences. This is due to the generally accurate information local politicians have about school performance in their municipality, which constitutes our third main finding. We show that this

reflects the high information level of local inhabitants, which improves citizens' ability to keep incumbents accountable and strengthens politicians' motives to acquire and use performance information (James and John 2007; Boyne et al. 2009; James and Moseley 2014).

Theoretical background

Since the NPM reforms of the 1980s, increasing stress has been put on performance in public sector organizations. This was reflected in a myriad of performance management schemes and systems brought forward and implemented by politicians across the political spectrum. Although the introduction of such schemes generally occurred with the aim of improving performance, recent evidence from a meta-analysis of 49 studies suggests that their impact on performance in public organizations may remain minimal (Gerrish 2016). Even so, this study also highlights that the impact of performance management systems increases substantially by the appropriate *management* of performance information. Benchmarking thereby plays a particularly important role since it "can generate a list of valid counterfactuals against which organizations can measure themselves" (Gerrish 2016: 62). Information about one's own performance relative to similar organizations thus may constitute an effective means to allow learning about what approaches are tied to better performance.

This is consistent with the theory of performance management, which argues that a central motivation behind the increasing stress on performance in public sector organizations is to help bureaucrats and elected officials make more informed decisions (Moynihan 2008; Nielsen and Baekgaard 2015). In fact, performance data are meant "to empower [decision-makers] to make informed decisions by presenting them with unambiguous information about the performance of institutions" (Baekgaard and Serritzlew 2016: 73). At the same time, performance measurement systems generally also lead to improved accountability (Moynihan 2008; Charbonneau and Van Ryzin 2015). For instance, the availability of performance data has been argued to improve citizens' ability to keep incumbent governments accountable at the polls (James and John 2007; Boyne et al. 2009; James and Moseley 2014). Knowing that citizens can and will use performance data to inform their decisions on election day provides politicians with a strong electoral incentive to bear in mind such data when developing their policy positions (Mueller 2003). An elected official's – or bureaucrat's – policy position thereby "indicates whether he or she favors (i.e., displays a positive attitude toward) or disfavors a

policy" (Andersen and Jakobsen 2017: 58). This leads to a first testable hypothesis about the link between public sector performance and politicians' reform preferences:

Hyp 1: Public sector performance affects politicians' reform preferences.

Clearly, the mere existence of performance data need not automatically imply that politicians also take this into consideration. Politicians should be informed about the level of public sector performance – and believe this information to be relevant and credible (Van Ryzin and Lavena 2013; James and Van Ryzin 2017; Nielsen and Moynihan 2017). Performance information is generally provided via some form of communication, which "surrounds every policy implemented by bureaucrats" (Andersen and Jakobsen 2017: 57). Yet, extensive research on public opinion formation highlights that individuals' attitudes are already susceptible to the communication of information. Communication about performance data first of all increases the salience of such information. This can be viewed as a 'framing' effect "emphasizing one subset of considerations rather than others" (Andersen and Jakobsen 2017: 57; Druckman et al. 2010). As such, it affects what politicians think about when forming their policy position, and is thereby likely to strengthen the link between public sector performance and reform preferences. Communication furthermore provides explicit or implicit cues – i.e. a "piece of information that enables individuals to make simplified evaluations without analyzing extensive information" (Andersen and Jakobsen 2017: 57; Druckman et al. 2010). Performance data reflect one such cue, and may act as a decision heuristic to reach a particular policy position. Again, this will work to strengthen the link between public sector performance and reform preferences. Our second hypothesis thus states that:

Hyp 2: Public sector performance affects politicians' reform preferences particularly when performance information is explicitly provided.

Performance information need not affect elected officials' policy position towards all types of reforms equally. The available menu of reform tools can range from restructuring allocated budgets (Nielsen and Baekgaard 2015; George et al. 2016) to affecting the governance characteristics of public services (including performance management and accountability; Figlio and Loeb 2011; Gerrish 2016) and changing the institutional design of service provision (including outsourcing and privatization; Boyne 1998; Geys and Sørensen 2016). Many types of policy reforms meet antagonism from key supporters, including government unions and

public employees (Pollitt 2006; Bonesrønning 2013). As indicated in the introduction, these political costs of reforms are likely to vary considerably with the characteristics of the reform proposal. This variation in political costs can be expected to induce heterogeneity in the effect of performance information on reform preferences across alternative types of (school) reforms.

More specifically, we argue that reforms with higher political costs require a closer tie to (poor) performance because such decisions require a stronger defense. The reason lies in the fact that negative information tends to have a stronger impact compared to positive information of equivalent magnitude (Tversky and Kahneman 1991). Recent (experimental) work on citizens' and politicians' responses to performance information likewise shows that especially information about negative performance induces stronger causal attribution of responsibility (Olsen 2015a; Nielsen and Moynihan 2017). Decision-makers are blamed for bad outcomes or unpopular decisions, but do not receive equal credit for good outcomes or popular decisions (James et al. 2016). This negativity bias provides decision-makers with an incentive to "follow a mini-max strategy and be more concerned with avoiding bad performances than with striving for excellence" (Olsen 2015a: 2; Hood and Dixon 2010). In our setting, negativity bias might reflect one potential micro-foundation for an increased need to defend politically costly reforms via performance information (which can point to an 'objective' need for such actions). As such, it is closely related to the notion that crises are good moments for reforms, because they increase the public's tolerance for remedial measures (Johnsen 2012; Asatryan et al. 2017). Drawing from the arguments on negativity bias, we expect the responsiveness of politicians' reform preferences to performance information depending on these reforms' political costs.

Hyp 3: Elected officials' reform preferences are most responsive to performance information for reform options with higher political costs.

Institutional context

Our analysis relies on multiple data sources covering Norwegian local governments. The institutional setting is a three-tier system comprising a central government, 19 county governments and 428 municipalities. National and local elections (for county and municipal governments) are held every fourth year, with an interval of two years between them. The local council is the main legislative body of the municipal government with full responsibility for all aspects of the municipality's activity, and consists of 11 to 85 members depending on

the size of the municipal population. It elects both a mayor and deputy mayor (who are responsible for chairing council meetings) and an executive board of minimum five members (which is responsible for the day-to-day running of the municipality). Unlike in a parliamentary system, the council – and not the executive board – is the key decision-making body, and councilors thus hold significant decision-making authority.

The local government sector in Norway accounts for approximately 20 percent of total employment, and is part of an integrated public sector where counties and municipalities are responsible for implementing national welfare policies, including the provision of primary and (lower) secondary schooling. Local governments own nearly all 3,100 primary and lower secondary schools (only 4.7% are private, non-profit schools, and about 3% of all pupils attend private schools). All schools – including the private schools – are completely tax-financed, and the education sector is subjected to extensive, central regulations. Central authorities define, for instance, a national core curriculum, minimum teacher qualifications and the amount of teaching offered. They also impose minimum safety- and hygienic standards for school buildings, and so on.

Nevertheless, and crucially, local authorities retain substantial autonomy to develop educational policies within the structure of this national framework. For example, the local council can allocate budgetary funds for a number of specific educational purposes (such as support to special-needs pupils), and has full flexibility to manage the school administration. Local governments may also develop their own strategies for recruiting the school's management team, decide on the size and location of schools, and implement self-defined performance indicators to oversee performance. As such, the local council has extensive powers to introduce school reforms whenever it deems such actions necessary or desirable.

Research design

We combine two main sources of data. First, we collected detailed registry data on school performance in Norwegian municipalities (as well as on social and economic variables characterizing the municipalities). Second, we gathered data on local council members' opinions about educational reforms in their municipality.

Registry data on school performance

The Norwegian Directorate for Education and Training (NDET) develops annual national tests designed to measure students' fundamental skills in reading, mathematics and English. Nearly all students participate in these national examinations, since only those entitled to special tuition may be granted exemption from the tests. Exemption rates therefore remain very low: 2% in mathematics, 2.3% in English and 3% in reading (Norwegian Directorate for Education and Training 2016). The assessment of all three examinations is standardized by computerization. The scores obtained by each pupil thus may be viewed as a measure of absolute performance (De Witte et al. 2014). NDET publishes average test scores at the municipal level, but pupils, parents, teachers and school management have access to pupils' individual test results. The national tests are not used as part of pupils' overall grade, but to customize teaching to their individual needs and improve the standard of education.

We use data on school performance measured at the eight grade (i.e. among 13-year old pupils), which are available from *www.skoleporten.no* (in Norwegian) for 388 municipalities. The data does not cover all 428 municipalities because results for municipalities with smaller student populations – where individual students might be identified – are not published. The scores are provided separately for mathematics, English and reading, and are made available on a standardized metric with an average of 50 and a student-level standard deviation of 10. Aggregated to the municipality-level, the data have an average close to 50 for all three tests, and a standard deviation at approximately 2.4.

For ease of comparison across municipalities, we calculate municipality-level percentage deviations from the national average. Similar measures of relative performance have been used in comparable contexts by, for instance, Charbonneau and Van Ryzin (2015), Nielsen and Baekgaard (2015), Barrows et al. (2016), Olsen (2016) and Nielsen and Moynihan (2017). These percentage deviations are measured as $MD_k = ((M_k - 50)*100)/50$, $ED_k = ((E_k - 50)*100)/50$, and $RD_k = ((R_k - 50)*100)/50$ (where M_k, E_k, and D_k reflect the average scores for the three test subjects across all pupils in municipality k). The overall test score for municipality k (TS_k) employed in our analysis is the average of these test-specific deviations (i.e. $TS_k = (MD_k + ED_k + RD_k)/3$), standardized with a standard deviation of 1 at the municipality level. Table A.2 in the Online Appendix displays descriptive statistics for this test score variable (before and after standardization), which highlights substantial cross-municipality variation in school performance (see also Fiva and Kirkebøen 2011).

Survey data and experimental design

Norway elected 10,621 council members in the local elections of 2015. We were able to collect e-mail addresses for 8,484 of these councilors, and distributed our survey to all these addresses in late May 2016. By mid-September 2016, we had obtained responses from 3,607 elected representatives across 417 municipalities (ranging from 1 to 31 answers per municipality). Since 592 of the collected e-mail addresses were inactive and 7,843 council members actually received the survey invitation, this yields an overall response rate of 46% (municipality-specific response rates vary from 3% to 73%). We conducted the survey with an explicit guarantee that all individual responses would be treated fully confidentially, which reduces potential self-serving biases in responses and benefits truthful replies (Heinemann et al. 2016).

The survey includes a set of background characteristics (including respondents' sex, age, education level, marital status, political position, and so on) as well as an experiment manipulating respondents' information about school performance. In the experiment, the 428 municipalities were randomly assigned to one of three information treatments using equal probabilities of assignment (between-subjects design), and all council members within a given municipality obtained the same treatment. A related design might have randomly allocated individual respondents rather than municipalities. We decided against this strategy as we were concerned that respondents in the same municipality might communicate (for instance, during council meetings or by phone), and thereby disrupt the assumption of independence across treatments. In the current context, the limitations of randomizing municipalities rather than individual respondents are small, since we can exploit a large sample of municipalities and council members. We can furthermore include relevant municipality-level controls in our analysis to counteract known sources of potential heterogeneity. Finally, as we show in table A.1. in the Online Appendix, balancing tests indicate that the three sets of councilors are very homogenous in terms of municipality-level and respondent-level characteristics. This confirms the success of the randomization process, and endorses the validity of the inferences drawn from our analysis.

Table 1 presents the performance information treatments as well as the order in which respondents were presented with specific questions or statements. Respondents in the control group were only asked to state their preferences regarding a number of school reform options aimed at improving school performance (more details below). In the other two groups, this reform question was preceded by another question about respondents' perception of school

performance in the 2015 national school tests. Answers were collected on a five-point scale ranging from "a lot poorer than the average in the country" to "much better than the average in the country" (with an option to reply "Do not know / Do not remember"; coded as missing values). While the 'perception' group was asked about school reform preferences immediately after reporting its performance perception, the 'treatment' group was additionally provided with the municipality's actual performance before seeing the reform question. Given our use of the real test scores, there is no deception in the experimental design. Clearly, however, most municipalities have more than one school, and pupils' performance may differ across these schools. Yet, we are not interested in the good or bad quality of any particular school. Rather, in line with publicly available performance information, our analysis – and thus also our question phrasing – focuses on school performance at the municipality level (see also Nielsen and Baekgaard 2015; Nielsen and Moynihan 2016, 2017).

_____

TABLE 1 ABOUT HERE

_____

The inclusion of the performance group in our design serves two main purposes. First, it provides useful information to investigate the accuracy of politicians' information about school performance (Meier et al. 2015; Nielsen and Moynihan 2016) – and thereby elucidates the likely impact of explicitly providing such information to our respondents in the treatment group. Second, school performance might commonly be thought of in very general terms. Providing performance information via test scores in the treatment group thus may lead two things to change at once relative to the control group: i.e. the provision of performance information and a more specific focus on test scores rather than general performance. The perception group focuses respondents' attention on test scores as a specific performance measure (which will increase its salience), and thus allows to disentangle both effects. This also allows testing whether focusing respondents' attention on test scores affects what politicians think about when faced with a set of reform proposals, which may have important practical implications regarding the framing of policy debates (Andersen and Hjortskov 2016; Andersen and Jakobsen 2017).

As is clear from table 1, respondents could select one of more reform options they consider especially important to improve school performance from a fixed-response menu. This menu only presented reform options under the direct control of municipal governments. More

encompassing reforms requiring national legislation – such as extending free school choice, increasing teacher wages or changing the length of compulsory basic education – were not included. Our list of options was also restricted to reforms that have been explicitly proposed and/or discussed by Norwegian local governments in recent years, although we provided the option for respondents to write in other reform proposals. Approximately 5% of respondents made use of this option, and offered a wide variety of answers. Since none of these achieved a high frequency, they were not retained in the final analysis. To develop our dependent variables, we recode answers to this question into a set of indicator variables equal to 1 when a specific reform is marked by the respondent, 0 otherwise. Table A.3 in the Online Appendix displays descriptive statistics for all eight reform options, as well as the 'no-reform' option. Only 6% of our respondents indicate that no reforms are necessary, while they on average marked 2.53 reform options with a standard deviation of 1.3 options. Explicitly accounting for the number of options marked by a given respondent in our analysis does not affect any of the findings reported below. The most commonly chosen reforms are linked to the recruitment of better-qualified teachers (41.6%), increased performance accountability for schools (40.8%) and improving assistance to weaker students (39.1%).

As all respondents received the reform items in the same order, one might be concerned about order effects. Yet, our respondents tend to have high education levels and empirical research suggests that such people are less prone to question wording, tone or order and response order effects (Narayan and Krosnick 1996; Geys and Heyndels 2003). Moreover, the most commonly provided responses are those that have attracted substantial debate in our Norwegian setting as (potential) avenues for improved school performance – rather than the first and last options presented in our list. Hence, respondents appear to have selected the reform options most relevant to the issue at hand, rather than be guided by the order of the options.

Finally, it is important to note that the performance information provided to the treatment group does not account for differences in the socio-economic composition of the municipalities' pupils (which tends to affect test scores). This might matter for our findings when politicians interpret the performance information based on knowledge about the socio-economic background of pupils. Recent work by Meier et al. (2015) suggests that middle managers in schools in Denmark and the US "do not provide sophisticated assessment of performance by giving their organization credit for the constraints it meets or discounting the resources it has" (Meier et al. 2015: 1084). We see little reason why politicians would be different (although this

is arguably an empirical question in need of further research). Still, even assuming that politicians make sophisticated assessments, this should lead them to discard at least part of the performance information provided. To the extent that such adjustments occur – which we cannot exclude – it would bias our findings towards representing a lower bound to the true effect of explicitly providing test scores.

Empirical analyses

School performance and policy reforms

Let R_{ik} denote individual-level responses to the reform indicators m1 to m9 (see table 1) for individual i in municipality k. We then estimate the following regression model:

$$R_{ik} = \propto_0 + \alpha^P + \alpha^T + \beta TS_k + Controls + \epsilon_{ik} \qquad (1)$$

where α^P and α^T are equal to 1 when the observation concerns the perception and treatment groups, respectively, and 0 otherwise (the control group is the excluded reference category). To test Hypothesis 1 we include a variable TS_k, which captures the test score in municipality k as defined above. If actual performance is unimportant for politicians' reform preferences, $\beta = 0$ (and non-zero otherwise). The set of control variables includes measures for (log) population size, the share of the municipal population aged 15 years or more with higher education, the share of the population in sparsely populated areas, the share of non-Western immigrants in the population, as well as politicians' age, gender, education, political experience, position as mayor or executive board member, and self-placement on an ideological left-right scale.

To assess whether the explicit provision of performance information matters (Hypothesis 2), we further differentiate TS_k into three parts. Let TS_k^C be school performance in municipality k assigned to the control group, and with values set to zero if the municipality has not been assigned to the control group. Similarly, TS_k^P and TS_k^T measure school performance in the perception and treatment groups. The regression model then becomes:

$$R_{ik} = \propto_0 + \alpha^P + \alpha^T + \beta^C TS_k^C + \beta^P TS_k^P + \beta^T TS_k^T + Controls + \epsilon_{ik} \qquad (2)$$

If performance is unimportant irrespective of the explicit provision of performance information, $\beta^T = \beta^P = \beta^C = 0$. If instead the explicit provision of information to politicians matters, we

should see a point estimate for β^T, which is statistically significantly different than that observed for β^P and β^C.

The results are summarized in table 2. Due to the binary nature of the dependent variables, we estimate logit models, and present coefficients in log odds form (with robust standard errors clustered at the municipality level in parentheses). Panel I assesses the overall effect of school performance, whereas Panel II estimates separate effects for the three information treatment groups. The F-test at the bottom of table 2 evaluates the equality of performance effects across the control, perception and treatment groups (that is, $\beta^T = \beta^P = \beta^C$).

_____

TABLE 2 ABOUT HERE

_____

A first observation in the top panel of table 2 is that, unsurprisingly, better school performance significantly increases the log odds of politicians desiring no reforms (column (1)). In terms of effect size, the point estimate suggests that a one-point increase in school performance (which is equivalent to just over one standard deviation in TS_k) increases the probability of marking the no-reform option with 22%, relative to the probability of not marking it (i.e. (Exp(0.197)–1)*100). Yet, school performance does not have similarly significant effects on all provided reform options. In effect, we only observe statistically significant effects for three out of eight reform options, such that is it important to account for heterogeneity across alternative types of (school) reforms.

A second observation concerns whether politicians' preferences respond to school performance depending on reforms' political costs (Hypothesis 3). As we have no precise measure of the political costs of the proposed reforms, we here rely on a simple comparison of estimated effects of school performance across the reform options. This is a valid approach since our information treatments are unlikely to influence perceived political costs of reforms, and only signal whether potential gains from any reform are high or low. Hence, our findings for the different information treatments are most likely to be driven by differences in political costs across the reforms. The results indicate that better (worse) performance significantly *reduces* (*increases*) the log odds of politicians supporting reforms aimed at recruiting better qualified teachers (column (6)) or introducing stricter school accountability (column (8)). This appears

reasonable given that such measures are explicitly aimed at increasing pupils' educational attainment (Figlio and Loeb 2011; Chetty et al. 2014). In terms of effect sizes, a one-point increase in school performance reduces the probability of desiring the recruitment of better qualified teachers with 17% and of desiring stricter school accountability with 14% (relative to the probability of not desiring these reforms). Interestingly, a one-point *increase* (decrease) in school performance also *increases* (decreases) the probability of desiring more support for weak students with 10% (relative to the probability of not desiring increased support). While initially surprising, this most likely reflects that good average performance of pupils in a municipality allows shifting attention to the remaining laggards, and explicitly aspire to improve their performance.

Turning to panel II in table 2, we find that the observations discussed above are largely driven by respondents in the 'treatment' group. That is, statistically significant effects in panel II tend to be concentrated in this group. Furthermore, pairwise tests of the coefficient estimates across the three information treatments indicate that we can formally reject equality of the coefficients for the treatment and control groups at the 90% confidence level for m3 (*Improve assistance to weak students*) and the 95% confidence level for m8 (*Keep schools accountable for results*). For m6 (*Recruit better qualified teachers*), the relevant p-value is 0.114. Still, the estimated coefficients are never statistically significantly different across the perception and treatment groups, nor in joint tests across all three information treatments (see the F-test at the bottom of table 2). This suggests that although explicitly providing information about school performance does have some effect on politicians' reform preferences, this influence remains statistically relatively weak. We return to this observation below.

From the results in table 2, it appears that reform options m6, m8 and – though to a lesser extent – m9 are most sensitive to performance information. A factor analysis – based on polychoric correlations to account for the binary nature of the data – confirms that these options are different from the others. Figure 1 presents the rotated factor loadings of the first two components extracted from this factor analysis. The factor loadings vary a lot more on the horizontal than the vertical axis, and only m6, m8 and m9 have positive loadings on the horizontal axis. As they thus appear to capture a clearly distinct subset of reforms, we combine our eight reform measures into two composite reform indicators. Specifically,

$$R_1 = \frac{m6 + m8 + m9}{3}$$

$$R_2 = \frac{m2 + m3 + m4 + m5 + m7}{5}$$

———————————

FIGURE 1 ABOUT HERE

———————————

Table A.3 in the Online Appendix displays descriptive statistics for the two composite reform indices (R_1, R_2), while Table A.4 shows a replication of table 2 using these reform indices as dependent variables. To illustrate that these two dimensions are theoretically meaningful, we exploit a question on spending allocations included in the survey: "Given the current level of municipal revenues, do you think the [education] program should receive a larger or a smaller share of total revenue, or do you think the current allocation is appropriate?" This question came *before* the survey experiment, and responses were coded as follows: 1 (larger share of revenues), 0 (current allocation is appropriate), and -1 (smaller share of revenues). The bivariate correlation between the R_1 index and preferences for education budget re-allocations is weakly negative (r=-0.07), whereas its correlation with R_2 is strong and positive (r=0.29). These correlations suggest that R_2 can interpreted as an index of resource-related reforms, whereas R_1 is best viewed as a set of governance-related reforms.

Actual and perceived school performance

Our results thus far indicate that the observed effects are predominantly driven by respondents in the treatment group. Yet, providing performance information appears to have an effect only relative to the control group, but not relative to the perception group. Asking respondents about their perception of school performance thus has (almost) the same effect as providing factual information. This strongly suggests that elected politicians already know school performance in their municipality and asking for their perception works to make this information salient to them. In this section, we test this proposition more directly linking data on politicians' performance perceptions to schools' actual performance from the registry data (for a similar approach, see Charbonneau and Van Ryzin 2012; Favero and Meier 2013). We only employ respondents in the perception group in this analysis. While respondents in the treatment group

likewise reported their perception of school performance (see table 1), they were subsequently shown the actual results. Since the survey software did not prevent respondents returning to previous screens, we cannot exclude that respondents in the treatment group adjusted their expressed perception of school performance after having seen the actual scores.

Figure 2 shows a municipality-level scatterplot of the relationship between actual school-level test scores (on the horizontal axis) and politicians' perceptions of these scores (on the vertical axis). Politicians' perceptions are averaged over all politicians from a given municipality answering our survey. Actual municipality-level test scores are measured as deviations from the national average. The bubble sizes are proportional to the number of observations in each of the 126 municipalities in the perception group, and the line represents a simple linear regression plot (with 95% confidence intervals). Figure 2 shows a substantively meaningful positive correlation between perceived and actual test scores, suggesting that local politicians are, on average, well informed about school performance in their municipality. This is consistent with Nielsen and Moynihan (2016: 2, italics in original), who conclude that Danish elected officials "appear to be knowledgeable enough about the *level* of organizational performance (…), but not so knowledgeable about the *causes* of organizational performance". Meier et al. (2015: 1093) likewise uncover a positive correlation between actual and perceived performance using data from US and Danish school principals, but argue that these correlations tend to remain modest. Even so, they also find that "Danish principals have slightly more accurate perceptions of performance that Texas principals" (Meier et al. 2015: 1093).[3]

_____

FIGURE 2 ABOUT HERE

_____

The analysis in Figure 3 suggests that this may reflect the force of local inhabitants' high information level. This figure depicts the relationship between municipality-level test scores in the eighth grade and user satisfaction with primary and secondary schools located in the relevant municipalities. The latter data were obtained from three large surveys conducted by the Agency for Public Management and eGovernment in 2010, 2013 and 2015 (see https://www.difi.no/om-difi/about-difi), and include 26,297 and 27,035 respondents answering a question about satisfaction with primary and secondary education, respectively. The question was: "We want to know your assessment of the services provided by [the following] agencies. If you do not have experience with a service, consider what you have seen, heard or read. How good or bad do you think these services are for those receiving / using it?" The responses were

recoded from their original seven-point scale (ranging from -3 to +3) into a scale from 0 to 100. The diagram on the left shows the relationship between primary school test scores and satisfaction with primary education, while the right diagram contains a placebo check linking primary school scores to satisfaction with secondary schools. While the first scatterplot indicates a strong and positive correlation, the second displays a weak and slightly negative correlation.

<center>_____</center>

<center>FIGURE 3 ABOUT HERE</center>

<center>_____</center>

Before concluding, we should rule out that our findings in this section are caused by respondents interrupting the (online) survey to look up school performance results. Since such action would only make sense in the two groups asked about their perceptions, we verify that response times did not vary across our three experimental groups (see column 12 in table A.1). We also replicated the analysis while excluding respondents taking more than 10 (i.e. the average response time), 15 or 20 minutes, which did not affect our results. One might furthermore worry that better informed politicians are more likely to answer the survey. Building on findings by Askim and Hanssen (2008) and Askim (2009) that education is significantly correlated with the likelihood to search and use performance information, we employ politicians' education level as a proxy for their information level. This shows that our results are robust when excluding respondents with higher education from the sample. Likewise, our results are robust to excluding respondents who required no or only few (i.e. 1, 2 or 3) reminders to answer the survey. (Note that if better informed politicians really self-select into answering (earlier), this should exclude the better informed respondents from the analysis.) Both results strongly suggest that our findings are not driven by potential self-selection of better informed respondents into our sample.

Discussion

Our findings indicate that school performance can have a meaningful effect on local politicians' preferences for reforms, but that such influence appears to be stronger for governance-related rather than resource-related reforms. A key lesson from our analysis thus is that focusing politicians' attention on school performance (measured via test scores) affects what politicians think about when faced with a set of reform proposals. As a result, they can be moved towards certain (types of) reform proposals by appropriately framing the discussion about school reform.

<center>16</center>

This is consistent with recent discussions about the impact of 'strategic communication' on frontline bureaucrats in Andersen and Jakobsen (2017), and can have important practical implications. Indeed, it suggest that strategically stressing certain types of information – such as absolute rather than relative performance statistics – might work to influence the stated policy preferences of political decision-makers.

A closely related lesson from our analysis is that performance statistics do not disappear into a black hole. Elected politicians in Norway are well informed about the relative performance of local schools. A similar positive relation between perceived and actual performance has recently also been observed among Danish elected officials (Nielsen and Moynihan 2016), although Meier et al. (2015) argue that such correlations remain modest among both US and Danish school principals. While the association between perceived and actual performance remains fairly modest also in our Norwegian setting, it is clear that politicians appear willing and able to adjust their policy preferences based on such feedback. As such, providing politicians with performance information is likely to have a measurable impact (see also Demaj 2017).

Finally, confirming earlier work by Charbonneau and Van Ryzin (2012), our results also indicate that the general public is well-informed about the performance of their local schools. This finding contributes to recent experimental work on citizens' use of performance information. Baekgaard and Serritzlew (2016) show that citizens' ability to interpret information about public sector performance depends on their prior beliefs. Olsen (2017) shows that episodic information (e.g., personal case stories) often crowds out quantitative performance data in citizens' recall of performance information. Our findings suggest that despite such likely influences on citizens' *interpretation* of performance data, they generally are able to avoid mis-attributions of good/bad performance across schools at different education levels. This clearly is crucial in keeping governments accountable by giving politicians an incentive to keep themselves informed as well (James 2011).

Conclusion

This article contributes to the recent literature on the relation between public performance use and reform preferences in the public sector in two ways (Salge 2011; Nielsen 2014; Nielsen and Baekgaard 2015; George et al. 2016; Nicholson-Crotty et al. 2017). First, we turn attention

to the reform preferences of political decision-makers rather than bureaucrats. This is important since elected politicians hold key decision-making power over structural reforms. Second, our analysis covers a broad spectrum of reform options. This allows characterizing heterogeneity in the sensitivity of various reforms to performance information, which we argue is consistent with differences in these reforms' political costs. Our findings suggest that school performance is most strongly related to preferences for (politically costly) governance-related reforms. Moreover, we show that local politicians in Norway are well informed about school performance, which can be linked to local citizens' high and specifically targeted information about outcomes in the education sector.

Future research should first of all assess the generalizability of these findings. One could argue that education is an atypical case since school results feature prominently in news media, and attracts considerable attention. In line with this argument, Askim (2007: 453) finds that performance information utilization is "higher among councilors working with elderly care, administrative affairs, and educational affairs". It would therefore be interesting to study how elected politicians respond to performance information in less salient policy fields, such as digital security, research quality or the efficiency of development aid. This could also help shed more light on the importance of citizens' information – which is likely to be lower for less salient policy fields – for politicians' information levels and accountability.

Future research would also benefit from analyzing natural or field experiments: e.g. data on politicians' preferences for (or decisions on) reforms before and after performance indicators were made available. This design would be particularly useful when the performance information is produced by higher-level governments or independent organizations (James and Van Ryzin 2017). One recent example in this direction is the experimental study of 57 actual legislators by Demaj (2017). He shows that providing politicians with performance information increases the likelihood of strong deviations from the status quo allocation.

A third avenue for further research concerns the particular importance that information about negative performance assessments appears to hold. While negativity bias in citizens' and politicians' perception of performance has recently gained renewed attention (Hood and Dixon 2010; Olsen 2015a,b; James et al. 2016; Nielsen and Moynihan 2017), the possible link between negative performance information and the "widely observed behavioural tendency of blame-avoidance in politics and public administration" (James, 2004; Hood 2007) has

remained understudied. As such, we have at best a limited understanding of the translation of negative performance information into various forms of blame avoidance by, and blame shifting between, politicians and bureaucrats. Recent experimental evidence, however, indicates that the option to shift blame can "constitute a strong motive for the delegation of a decision right" (Bartling and Fischbacher 2012: 67). This at least suggests that delegation of authority and the outsourcing of public policies may gain in prominence when politicians believe that negative outcomes are more likely to occur – and the option to shift blame becomes more valuable.

Finally, existing studies on the impact of performance information mostly focus on either politicians or street-level bureaucrats. No analyses as yet exist addressing the possible inter-relation between administrative and political responses to performance data. One hypothesis might be that government agencies only take action to improve on poor performance when they believe politicians otherwise may take harsh(er) steps affecting employees negatively. Alternatively, government agencies might respond independently of expected political reactions, particularly when staff members think that the indicators offer a fair representation of agency performance (e.g., when they derive from independent organizations). Empirical studies addressing these hypotheses would be very valuable to gain a more complete understanding of policy-makers' responses to performance information.

ENDNOTES:

1. Nielsen and Moynihan (2016a,b) use survey experiments to look at how Danish local politicians attribute bureaucratic responsibility for performance, but do not link this to reform preferences. Baekgaard et al. (2017) use survey experiments among a similar subject pool to assess whether performance information is more likely to be ignored when it is at odds with politicians' prior attitudes.

2. We performed a content analysis of the Norwegian media archive – Atext – over the period 2009-2016. Using the search string "school closures", we see a peak in the number of news stories during the final rounds of municipalities' budget preparations in the fourth quarter of the year (details upon request).

3. The analysis in Online Appendix A.4 confirms these results using a regression-based framework with extensive control variables.

References

Andersen, Simon Calmar and Morten Hjortskov. 2016. Cognitive Biases in Performance Evaluations. *Journal of Public Administration Research and Theory* 26(4): 647-662.

Andersen, Simon Calmar and Morten Jakobsen. 2017. Policy Positions of Bureaucrats at the Front Lines: Are they Susceptible to Strategic Communication? *Public Administration Review* 77(1): 57-66.

Andersen, Simon Calmar and Donald P. Moynihan. 2016. Bureaucratic Investments in Expertise: Evidence from a Randomized Controlled Field Trial. *Journal of Politics* 87(4): 1032-1044.

Asatryan, Zareh, Friedrich Heinemann and Hans Pitlik. 2017. Reforming the Public Administration: The Role of Crisis and the Power of Bureaucracy. *European Journal of Political Economy* forthcoming.

Askim, Jostein. 2007. How do Politicians use Performance Information? An Analysis of the Norwegian Local Government Experience. *International Review of Administrative Sciences* 73(3): 453-472.

Askim, Jostein and Gro Sandkjaer Hanssen. 2008. Councillors' Receipt and Use of Citizen Input: Experience from Norwegian Local Government. *Public Administration* 86(2): 387-409.

Askim, Jostein. 2009. The Demand Side of Performance Measurement: Explaining Councillors' Utilization of Performance Information in Policymaking. *International Public Management Journal* 12(1): 24-47.

Askim, Jostain, Åge Johnsen and Knut-Andreas Christophersen. 2008. Factors behind Organizational Learning from Benchmarking: Experiences from Norwegian Municipal Benchmarking Networks. *Journal of Public Administration Research and Theory* 18: 297-320.

Baekgaard, Martin and Søren Serritzlew. 2016. Interpreting Performance Information: Motivated Reasoning or Unbiased Comprehension. *Public Administration Review* 76(1): 73-82.

Baekgaard, Martin, Julian Christensen, Casper M. Dahlmann, Asbjørn Mathiasen and Niels B.G. Petersen. 2017. The Role of Evidence in Politics: Motivated Reasoning and Persuasion among Politicians. *British Journal of Political Science*, forthcoming.

Barrows, Samuel, Michael Henderson, Paul E. Peterson, and Martin R. West. 2016. Relative Performance Information and Perceptions of Public Service Quality: Evidence from American School Districts. *Journal of Public Administration Research and Theory*, 26, 571-583.

Bartling, Björn, and Urs Fischbacher. 2012. Shifting the Blame: On Delegation and Responsibility. *Review of Economic Studies* 79(1): 67-87.

Bjørnholt, Bente, Martin Bækgaard and Kurt Houlberg. 2016. Does Fiscal Austerity Affect Political Decision-Makers' Use and Perception of Performance Information? *Public Performance & Management Review* 39(3): 560-580.

Bonesrønning, Hans. 2013. Public Employees and Public Sector Reform Implementation. *Public Choice* 156: 309-327.

Boyne, George. 1998. The Determinants of Variations in Local Service Contracting: Garbage in, Garbage out? *Urban Affairs Review* 34(1): 150-163.

Boyne, George, Oliver James, Peter John and Nicolai Petrovsky. 2009. Democracy and Government Performance: Holding Incumbents Accountable in English Local Governments. *Journal of Politics* 71(4): 1273-1284.

Charbonneau, Étienne and Gregg 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 Van Ryzin. 2015. Benchmarks and Citizen Judgments of Local Government Performance: Findings from a Survey Experiment. *Public Management Review* 17(2): 288-304.

Crozier, Michel. 1964. *The Bureaucratic Phenomenon*. Chicago: University of Chicago Press.

Chetty, Raj, John N. Friedman and Jonah E. Rockoff. 2014. Measuring the Impacts of Teachers I: Evaluating Bias in Teacher Value-Added Estimates. *American Economic Review* 104(9): 2593-2632.

Demaj, Labinot. 2017. What Can Performance Information Do to Legislators? A Budget-Decision Experiment with Legislators. *Public Administration Review* 77: 366-379.

De Witte, Kristof, Benny Geys and Catharina Solondz. 2014. Public Expenditures, Educational Outcomes and Grade Inflation: Theory and Evidence from a Policy Intervention in the Netherlands, *Economics of Education Review* 40: 152-166.

Druckman, James N., Cari Lynn Hennessy, Kristi St. Charles, and Jonathan Webber. 2010. Competing Rhetoric over Time: Frames versus Cues. *Journal of Politics* 72(1): 136-148.

Favero, Nathan and Kanneth J. Meier. 2013. Evaluating Urban Public Schools: Parents, Teachers, and State Assessments. *Public Administration Review* 73: 401-412.

Figlio, David and Susanna Loeb. 2011. School Accountability. In *Handbook of Economics of Education (Volume 3)*, edited by Eric Hanushek, Stephen Machin, and Ludger Woessmann. Amsterdam: North Holland.

Fiva, Jon and Lars Kirkebøen. 2011. Information Shocks and the Dynamics of the Housing Market. *Scandinavian Journal of Economics* 113: 525-552.

George, Bert, Sebastian Desmidt, Poul Nielsen and Martin Baekgaard. 2016. Rational Planning and Politicians' Preferences for Spending and Reform: Replication and Extension of a Survey Experiment. *Public Management Review* forthcoming.

Gerrish, Ed. 2016. The Impact of Performance Management on Performance in Public Organizations: A Meta-Analysis. *Public Administration Review* 76(1): 48-66.

Geys, Benny and Bruno Heyndels. 2003. The Influence of 'Cognitive Sophistication' on Ballot Layout Effects. *Acta Politica*, 38(4): 295-311.

Geys, Benny and Rune Sørensen. 2016. Revenue Scarcity and Government Outsourcing: Evidence from Norwegian Local Governments. *Public Administration* 94(3): 769-788.

Heinemann, Friedrich, Eckhard Janeba, Christoph Schröder and Frank Streif. 2016. Fiscal Rules and Compliance Expectations: Evidence for the German Debt Brake. *Journal of Public Economics* 142: 11-23.

Hood, Christopher. 2007. What Happens when Transparency meets Blame-avoidance? *Public Management Review* 9(2): 191-210.

Hood, Christopher and Ruth Dixon. 2010. The political payoff from performance target systems: No-brainer or no-gainer? *Journal of Public Administration Research and Theory* 20(2): 281-298.

James, Oliver. 2004. The UK Core Executive's Use of Public Service Agreements as a Tool of Governance. *Public Administration* 82(2): 397-419.

James, Oliver. 2011. 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 and Peter John. 2007. Public Management at the Ballot Box: Performance Information and Electoral Support for Incumbent English Local Governments. *Journal of Public Administration Research and Theory* 17(4): 567-580.

James, Oliver and Alice Moseley. 2014. Does Performance Information about Public Services Affect Citizens' Perceptions, Satisfaction and Voice Behaviour? Field Experiments with Absolute and Relative Performance Information. *Public Administration* 92(2): 493-511.

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: 83-93.

James, Oliver and Gregg G. Van Ryzin. 2017. Incredibly Good Performance: An Experimental Study of Source and Level Effects on the Credibility of Government. *American Review of Public Administration* 47(1): 23-35.

Johnsen, Åge. 2012. Why Does Poor Performance get so much Attention in Public Policy? *Financial Accountability & Management* 28(2): 121-142.

Kroll, Alexander. 2015. Drivers of Performance Information Use: Systematic Literature Review and Directions for Future Research. *Public Performance and Management Review*, 38, 3, 459-486.

Meier, Kenneth J., Søren C. Winter, Laurence J. O'Toole, Nathan Favero and Simon Calmar Andersen. 2015. The Validity of Subjective Performance Measures: School Principals in Texas and Denmark. *Public Administration* 93(4): 1084-1101.

Moynihan, Donald P. 2008. *The Dynamics of Performance Management: Constructing Information and Reform*. Washington, DC: Georgetown University Press.

Moynihan, Donald P. and Sanjay Pandey. 2010. The Big Question for Performance Management: Why Do Managers Use Performance Information? *Journal of Public Administration Research and Theory* 20: 849-866.

Mueller, Dennis. 2003. *Public Choice III*. Cambridge: Cambridge University Press.

Narayan, S. and Jon Krosnick 1996. Education Moderates Some Response Effects in Attitude Measurement. *Public Opinion Quarterly*, 60: 58-88.

Nicholson-Crotty, Sean, Jill Nicholson-Crotty and Sergio Fernandez. 2017. Performance and Management in the Public Sector: Testing a Model of Relative Risk Aversion. *Public Administration Review* 77(4).

Nielsen, Poul. 2014. Learning from Performance Feedback: Performance Information, Aspiration Levels, and Managerial Priorities. *Public Administration* 92: 142-160.

Nielsen, Poul and Martin Baekgaard. 2015. Performance Information, Blame Avoidance, and Politicians' Attitudes to Spending and Reform: Evidence from an Experiment. *Journal of Public Administration Research and Theory* 25(2): 545-569.

Nielsen, Poul and Donald P. Moynihan. 2016. Romanticizing Bureaucratic Leadership? The Politics of how Elected Officials Attribute Responsibility for Performance. *Governance*, forthcoming.

Nielsen, Poul and Donald P. Moynihan. 2017. How Do Politicians Attribute Bureaucratic Responsibility for Performance? Negativity Bias and Interest Group Advocacy. *Journal of Public Administration Research and Theory*, 27(2): 269-283.

Norwegian Directorate for Education and Training. 2016. Metodegrunnlag for nasjonale prøver. *Rapport*. [Methodological basis of national tests. *Report*] Oslo.

Olsen, Asmus. 2015a. Negative Performance Information Causes Asymmetrical Evaluations and Elicits Strong Responsibility Attributions. Paper presented at 2015 APSA Meeting, *mimeo*.

Olsen, Asmus. 2015b. Citizen (Dis)satisfaction: An Experimental Equivalence Framing Study. *Public Administration Review*, 75: 469–478.

Olsen, Asmus. 2016. How Social and Historical Reference Points Affect Citizens' Performance Evaluations. *Journal of Public Administration Research and Theory*, forthcoming.

Olsen, Asmus. 2017. Human Interest or Hard Numbers? Experiments on Citizens' Selection, Exposure, and Recall of Performance Information. *Public Administration Review*, 77(3): 408-420.

Pollitt, Christopher. 2006. Performance Management in Practice: A Comparative Study of Executive Agencies. *Journal of Public Administration Research and Theory* 16(1): 25-44

Salge, Oliver. 2011. A Behavioral Model of Innovative Search: Evidence from Public Hospital Services. *Journal of Public Administration Research and Theory* 21(1): 181-210.

Ter Bogt, Henk J. 2004. Politicians in Search of Performance Information? Survey Research on Dutch Aldermen's Use of Performance Information. *Financial Accountability and Management* 20(3): 221-252.

Tversky, Amos and Daniel Kahneman. 1991. Loss Aversion in Riskless Choice: A Preference-dependent Model. *Quarterly Journal of Economics* 106(4): 1039-1061.

Van Ryzin, Gregg G. and Cecelia F. Lavena. 2013. The Credibility of Government Performance Reporting. *Public Performance & Management Review* 37(1): 87-103.

Figure 1. Factor analysis. Rotated factor loadings
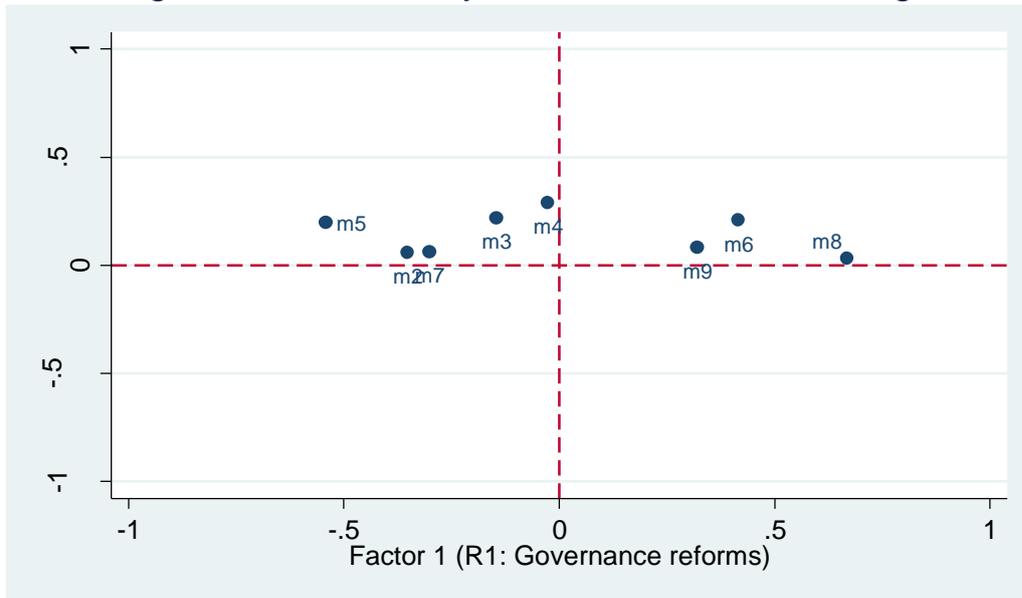
Notes. The factor analysis assumes two components, and estimates are based on polychoric correlations.
Legend. m2:Free lunch, m3:Improve assistance to weak students, m4:Improve pre-school services, m5:Employ more teachers, m6:
Recruit better qualified teachers, m7:Allow teachers more discretion, m8: Keep schools accountable for results, m9:Close
schools with small teaching staffs. The labels 'Governance reforms' and 'Resource reforms' indicate our interpretations
of the two factors.

Figure 2. Perceived and actual test scores
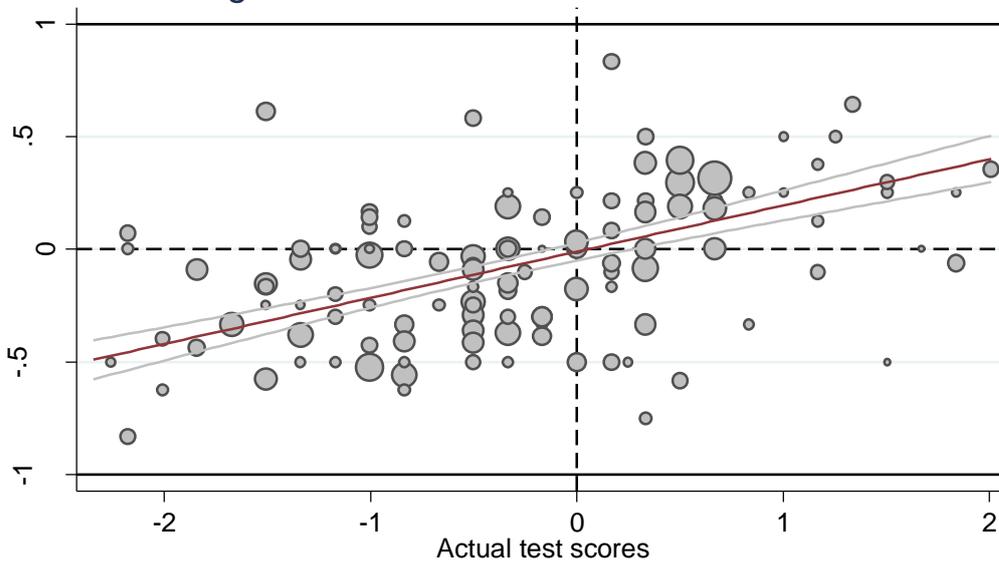
Note. The test score perceptions are measured as municipality-level averages. The individual responses are coded -1
(much lower than average), -0.5 (lower than average) 0 (average), 0.5 (higher than average), and 1 (much higher than average).
The test scores are measured as deviations for the national average, and is the mean scores obtained in reading, english and
mathematics. The test score has been calculated with a standard deviation of 1.
The bubble sizes are proportional to the square root of number of respondents per municipality.
The graph includes a linear regression line and a corresponding 95% confidence interval.

Figure 3. School satisfaction and test scores
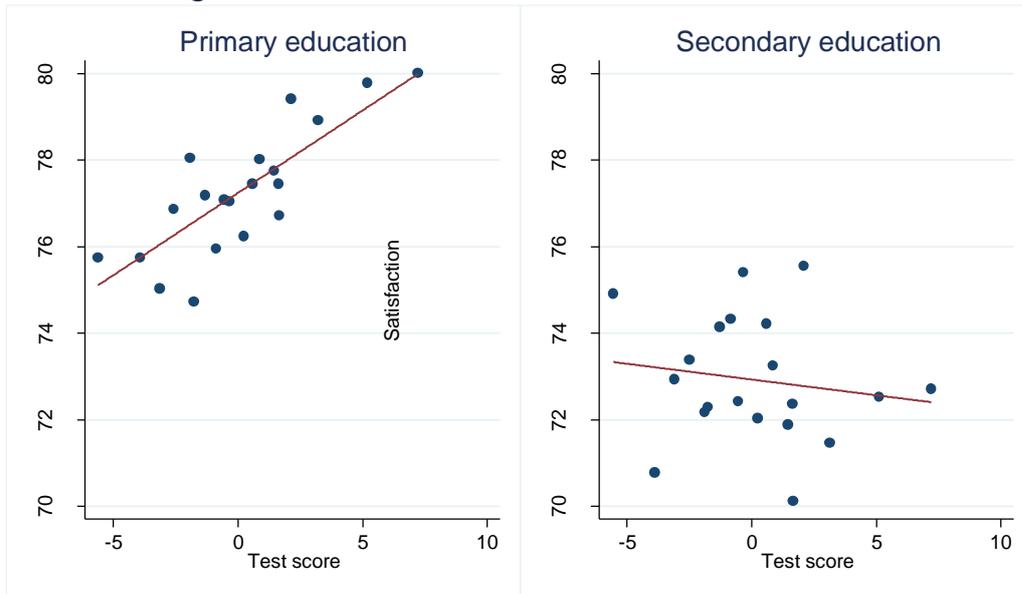
Note. The graph is a binned scatterplot with satisfaction with schooling and average test scores in mathematics, English and reading in the eighth grade. Satisfaction is measured on a scale from 0 (very poor) to 100 (very good). The diagram is a partial scatterplot with controls for municipal population size (log-scale).

Table 1: Experimental treatments

Control Group	Perception Group	Treatment Group
	In 2015, national tests were carried out in the eighth grade in mathematics, reading, and English. Do you remember how well pupils in your municipality fared in these tests?	
		Expressed as the percentage deviation from the national average, your municipality obtained the following results in the 2015 national tests. *English: ... %* *Reading: ... %* *Mathematics: ... %*

Which of these actions do you think are especially important to implement in order to improve school performance in your municipality? (Feel free to provide multiple answers)

    m1: No special measures are necessary in our municipality.

    m2: The municipality should offer free school lunches.

    m3: The municipality should improve services for disadvantaged pupils, including more special education.

    m4: The municipality should strengthen daycare services so that children are better prepared for school.

    m5: The municipality should hire more teachers.

    m6: The municipality should recruit better-qualified teachers.

    m7: The municipality should give teachers more freedom to organize their teaching.

    m8: The municipality should set clearer goals for schools' learning outcomes and principals should be held accountable if targets are not met.

    m9: The municipality should close down schools with smaller teaching staff.

Table 2. School test scores and policy responses measured by individual survey items

Variables	(1) m1	(2) m2	(3) m3	(4) m4	(5) m5	(6) m6	(7) m7	(8) m8	(9) m9
				Panel I: Overall test score effect					
TS	0.197**	0.042	0.095*	0.041	-0.073	-0.186***	-0.033	-0.133**	-0.107
	(0.095)	(0.058)	(0.055)	(0.060)	(0.060)	(0.054)	(0.049)	(0.053)	(0.083)
Perception (=1)	0.060	0.065	-0.145	0.014	-0.082	-0.123	0.090	-0.176*	-0.251
	(0.232)	(0.120)	(0.111)	(0.121)	(0.119)	(0.103)	(0.093)	(0.101)	(0.157)
Treatment (=1)	-0.004	-0.054	-0.093	-0.062	-0.227*	-0.036	0.113	-0.241**	-0.121
	(0.227)	(0.116)	(0.110)	(0.123)	(0.122)	(0.097)	(0.096)	(0.102)	(0.153)
Obs.	3,495	3,495	3,495	3,495	3,495	3,495	3,495	3,495	3,495
Pseudo-R^2	0.084	0.083	0.038	0.023	0.127	0.055	0.027	0.112	0.045
				Panel II: Differentiated test score effects					
TS Control	0.195	0.059	0.007	-0.051	-0.171	-0.112	0.017	-0.032	-0.159
	(0.149)	(0.081)	(0.079)	(0.098)	(0.105)	(0.090)	(0.082)	(0.071)	(0.120)
TS Perception	0.047	0.061	0.117	0.085	0.002	-0.186**	-0.049	-0.146*	-0.112
	(0.157)	(0.090)	(0.091)	(0.083)	(0.086)	(0.077)	(0.082)	(0.087)	(0.143)
TS Treatment	0.442**	-0.008	0.198**	0.115	-0.028	-0.293***	-0.077	-0.269***	-0.021
	(0.205)	(0.109)	(0.083)	(0.103)	(0.125)	(0.080)	(0.078)	(0.099)	(0.141)
Perception (=1)	0.018	0.065	-0.109	0.059	-0.021	-0.149	0.069	-0.216**	-0.231
	(0.241)	(0.131)	(0.109)	(0.124)	(0.124)	(0.114)	(0.095)	(0.106)	(0.161)
Treatment (=1)	0.065	-0.080	-0.025	-0.003	-0.173	-0.109	0.080	-0.339***	-0.059
	(0.226)	(0.121)	(0.110)	(0.133)	(0.139)	(0.103)	(0.100)	(0.109)	(0.176)
Obs.	3,495	3,495	3,495	3,495	3,495	3,495	3,495	3,495	3,495
Pseudo-R^2	0.086	0.083	0.039	0.024	0.128	0.056	0.027	0.13	0.045
F-test	2.25	0.34	3.22	1.77	1.74	2.63	0.79	4.29	0.62
Prob > F	0.325	0.843	0.200	0.413	0.419	0.269	0.674	0.117	0.733

Notes. The response variables are preferences for school reform as measured by the eight survey items used to measure reform preferences, c.f. Table 1. The school test scores and response variables are normalized with a standard deviation of 1. The models are logit models, and we present coefficients as odds ratios (robust standard errors, presented in parentheses, are clustered at the municipality level). The regressions include a set of controls for (log) population size, share of population with higher education, share of population in sparsely populated areas, share of non-Western immigrants in population and respondents' age, gender, education, political experience, position as mayor or executive board member, and self-placement on an ideological left-right scale. Panel I assesses the overall effect of test scores, whereas Panel II estimates separate test score effects for the three information treatment groups. In this case, the F-test tests the equality of test score effects across the control, perception and treatment groups (that is, $\beta^T = \beta^P = \beta^C$). Significance levels: *** p<0.01, ** p<0.05, * p<0.1.

Online Appendix to

Never Change a Winning Policy?
Public Sector Performance and Politicians' Preferences for Reforms

A.1. Balancing tests

Table A.1 provides the findings from a set of balancing tests, which compare a number of variables across councilors (in columns (1) through (6)) and respondents (in columns (7) through (9)) in the three information treatment groups. The results indicate that response rates, school test scores, and the municipal population size, settlement pattern, education level and number of immigrants are very similar across councilors in the three information groups. The same also holds for individual respondents' position on a left-right self-placement scale, gender, (vice-)mayor post, executive board membership, and education level. We should note that individuals' age and political experience as party members were asked in ten- and five-year categories, respectively. While this makes it more difficult to display in table A.1, a Pearson Chi^2 test indicates that also these characteristics show very similar distributions across councilors in the three information groups. For politicians' age, the test statistic is 9.53 (p=0.657; 12 degrees of freedom), whereas for their political experience the test statistic is 2.65 (p=0.618; 4 degrees of freedom).

_____

TABLE A.1

ABOUT HERE

_____

As an additional test, we also estimate a multinomial regression model with the three information treatment groups as response variables. The regression includes all variables included in Table A.1 as explanatory variables. A test of the joint significance of all these variables provides a Chi^2 test statistic of 13.74. With 28 degrees of freedom and 3,495 observations, this gives a significance probability of 0.989. Moreover, none of the included variables is individually significant at conventional levels. Overall, therefore, both sets of tests strongly indicate that the randomization within our research design has produced three homogenous subsamples.

Table A.1. Balancing test results

Variables	(1) Response	(2) Test score	(3) Population (log)	(4) Higher education	(5) Non-Western immigrants	(6) Settlement pattern	(7) Left-right placement	(8) Gender	(9) Board member	(10) Mayor	(11) University degree	(12) Response Time
Control	0.427	-0.331	8.989	0.232	0.041	0.400	5.298	0.626	0.233	0.116	0.649	9.976
group	(0.010)	(0.083)	(0.126)	(0.007)	(0.003)	(0.025)	(0.059)	(0.012)	(0.009)	(0.008)	(0.016)	(0.216)
Perception	0.425	-0.303	8.977	0.231	0.044	0.397	5.239	0.622	0.227	0.113	0.654	9.9976
group	(0.011)	(0.088)	(0.137)	(0.007)	(0.004)	(0.027)	(0.061)	(0.014)	(0.012)	(0.010)	(0.016)	(0.232)
Treatment	0.429	-0.467	9.123	0.226	0.043	0.385	5.386	0.630	0.253	0.111	0.663	9.726
group	(0.011)	(0.087)	(0.122)	(0.007)	(0.003)	(0.025)	(0.067)	(0.014)	(0.012)	(0.008)	(0.018)	(0.176)
Obs.	8,434	8,105	8,419	8,419	8,419	8,419	3,602	3,602	3,602	3,602	3,602	3,583
R-squared	0.427	0.143	0.982	0.930	0.705	0.680	0.864	0.626	0.238	0.120	0.655	0.697
F-test	0.04	1.03	0.41	0.18	0.21	0.10	1.31	0.08	1.23	1.34	0.17	0.56
Prob > F	0.962	0.360	0.661	0.833	0.811	0.908	0.270	0.923	0.294	0.262	0.848	0.572

Notes. The F-test assesses homogeneity across the three information groups of council members. 'Response' is an indicator variable equal to 1 if the councilor provided a complete answer to our survey, 'test score' is the TS variable defined in the main text, 'higher education' is the share of the municipal population aged 15 years of more with higher education, 'settlement patterns' is the share of the population living in sparsely populated areas. Then, 'Left-right placement' is his/her placement on an 11-point scale from Left (0) to Right (10), while 'Gender', 'Board member', 'Mayor', and 'University degree' are indicator variables equal to 1 for men, executive board members, (vice-)mayors, and respondents with a university degree, respectively. Finally, 'Response time' is the number of minutes taken to complete the survey. Robust standard errors clustered at the municipality level in parentheses.

A.2. Descriptive statistics

Table A.2: Descriptive statistics. Municipal-level variables

VARIABLES	(1) N	(2) mean	(3) sd
Test Score	396	-1.720	3.984
Test score, standardized	396	-0.432	1.000
Population size (in 2015)	428	12,070	37,405
Share of population living in dispersed settlements	428	0.482	0.279
Share with higher education	428	0.213	0.0558

Table A.3: Descriptive statistics. Reform preferences

VARIABLES	(1) N	(2) mean	(3) sd
m1 - No specific measures required	3,663	0.058	0.229
m2 - Free lunch	3,663	0.198	0.399
m3 - Improve assistance to weak students	3,663	0.391	0.488
m4 - Improve pre-school services	3,663	0.257	0.437
m5 - Employ more teachers	3,663	0.361	0.480
m6 - Recruit better qualified teachers	3,663	0.416	0.493
m7 - Allow teachers more discretion	3,663	0.316	0.465
m8 - Keep schools accountable for results	3,663	0.408	0.492
m9 - Close schools with small teaching staffs	3,663	0.184	0.388
Number of reform initiatives mentioned	3,663	2.531	1.298
Governance index (R1)	3,663	0.336	0.304
Resource index (R2)	3,663	0.305	0.226

A.3. Regression results using the reform indices

Table A.4 – which has the same format as table 2 in the main text – presents regression estimates using these two reform indices as independent variables. For both reform types, we provide results with and without control variables to illustrate that our inferences are not sensitive to including a relevant set of controls. Overall, the results in table A.4 illustrate that school performance has a negative impact on preferences for governance-related reforms, suggesting that poor (good) performance motivates (discourages) politicians to favor a new governance regime. Panel II shows that the estimated effect sizes are again smaller in absolute values in the control and perception groups compared to the treatment group. A pairwise comparison on the equality of the estimated coefficients for the control and treatment groups reaches a p-value of 0.115. Hence, accurate information does appear to have some effect, although the F-tests point out that any differences across the three groups are not statistically significant. Columns (3) and (4) indicate that test scores have no significant relation to preferences for resource-related reforms once relevant control variables are included.[1]

---

[1] One potential explanation for this non-finding may be the high level of resources already allocated to the education sector in the average Norwegian municipality (i.e. 21% of total spending in 2013), which may make politicians less inclined to opt for a further increase in resources. We return to this in the concluding discussion.

Table A.4. School test scores and policy response

Variables	(1) R_1	(2) R_1	(3) R_2	(4) R_2
	Panel I: Overall test score effect			
TS	-0.107***	-0.093***	0.070***	0.013
	(0.021)	(0.023)	(0.023)	(0.021)
Perception (=1)	-0.104**	-0.109**	0.008	-0.014
	(0.048)	(0.046)	(0.061)	(0.047)
Treatment (=1)	-0.053	-0.078*	-0.047	-0.051
	(0.051)	(0.046)	(0.056)	(0.047)
Controls	NO	YES	NO	YES
Obs.	3,495	3,495	3,495	3,495
R-squared	0.011	0.163	0.005	0.164
	Panel II: Differentiated test score effects			
TS Control	-0.077**	-0.062*	0.008	-0.025
	(0.038)	(0.036)	(0.040)	(0.035)
TS Perception	-0.097***	-0.094**	0.083*	0.026
	(0.032)	(0.036)	(0.042)	(0.032)
TS Treatment	-0.153***	-0.136***	0.131***	0.037
	(0.037)	(0.035)	(0.036)	(0.039)
Perception (=1)	-0.111**	-0.120**	0.033	0.007
	(0.051)	(0.049)	(0.065)	(0.046)
Treatment (=1)	-0.082	-0.107**	-0.001	-0.028
	(0.054)	(0.048)	(0.059)	(0.050)
Controls	NO	YES	NO	YES
Obs.	3,495	3,495	3,495	3,495
R-squared	0.012	0.164	0.007	0.165
F-test	1.110	1.262	2.651	1.105
Prob > F	0.331	0.284	0.072*	0.332

Notes. The response variables are preferences for school reform as measured by the indexes R_1 and R_2. The school test scores and response variables are normalized with a standard deviation of 1. Columns (2) and (4) include a set of control variables for (log) population size, share of population with higher education, share of population in sparsely populated areas, share of non-Western immigrants in population and respondents' age, gender, education, political experience, position as mayor or executive board member, and self-placement on an ideological left-right scale. The model are linear regression models and standard errors (presented in parentheses) are robust standard errors clustered at the municipality level. Panel I assesses the overall effect of test scores, whereas Panel II estimates separate test score effects for the three information treatment groups. In this case, the F-test tests the equality of test score effects across the control, perception and treatment groups (that is, $\beta^T = \beta^P = \beta^C$). Significance levels: *** p<0.01, ** p<0.05, * p<0.1

A.4. School satisfaction and test scores

In table A.5, we assess the relation between actual and perceived school performance using a regression model including additional individual- and municipality-level controls as well as municipality-level random effects (standard errors are clustered at the municipality level). The bivariate regression result in column (1) corroborates the pattern observed in figure 2 in the main text. Including individual- and municipality-level controls in column (2) does not change this relationship. Column (3) adds a quadratic term to assess whether the relationship might be non-linear, which appears not to be the case. These supplementary regressions thus lend further support to the hypothesis that elected politicians are on average well informed about school test scores.

Table A.5. Perceived and actual test scores

	(1)	(2)	(3)
Test score	0.121***	0.101***	0.110***
	(0.030)	(0.024)	(0.025)
Test score, squared			0.008
			(0.016)
Controls	NO	YES	YES
Observations	953	953	953
Number of municipalities	126	126	126
Sigma_u	0.262	0.202	0.203
Sigma_e	0.330	0.330	0.330

Notes. The response variable is a politician's perception of test scores within his municipality on an ordinal scale with values -1, -0.5, 0, 0.5 and 1 (see main text for details). The models are random-effects models using individual-level data, where Sigma_u denotes the municipality-component, and Sigma_e denotes the idiosyncratic component. The set of control variables includes (log) population size, share of population with higher education, share of population in sparsely populated areas, share of non-Western immigrants in population and respondents' age, gender, education, political experience, position as mayor or executive board member, and self-placement on an ideological left-right scale. Robust standard errors clustered at the municipality level in parentheses. Significance levels: *** $p<0.001$, ** $p<0.01$, * $p<0.05$.