How Does Public Service Motivation Among Teachers Affect Student Performance in Schools?

Andersen, Lotte Bøgh; Heinesen, Eskil; Holm Pedersen, Lene

Document Version
Accepted author manuscript

Published in:
Journal of Public Administration Research and Theory

DOI:
10.1093/jopart/mut082

Publication date:
2014

License
Unspecified

Citation for published version (APA):
Andersen, L. B., Heinesen, E., & Holm Pedersen, L. (2014). How Does Public Service Motivation Among Teachers Affect Student Performance in Schools? *Journal of Public Administration Research and Theory*, *24*(3), 651-671. https://doi.org/10.1093/jopart/mut082

How Does Public Service Motivation Among Teachers Affect Student Performance in Schools?

Lotte Bøgh Andersen, Eskil Heinesen, and Lene Holm Pedersen

How does public service motivation among teachers affect student performance in schools?

Lotte Bøgh Andersen
Aarhus University & Danish Institute for Local and Regional Government Research
Bartholins Allé 7
DK-8000 Aarhus C.
Denmark
(lotte@ps.au.dk),

Eskil Heinesen
Rockwool Foundation Research Unit
Sølvgade 10
DK-1307 Copenhagen K.
Denmark
(esh@rff.dk)

Lene Holm Pedersen
CBS – Copenhagen Business School & Danish Institute for Local and Regional Government Research
Steen Blichers Vej 22
DK-2000 Frederiksberg
Denmark
(lhp.dbp@cbs.dk)

How does public service motivation among teachers affect student performance in schools?

Abstract

The literature expects public service motivation (PSM) to affect performance, but most of the existing studies of this relationship use subjective performance data and focus on output rather than outcome. This article investigates the association between PSM and the performance of Danish teachers using an objective outcome measure (the students' academic performance in their final examinations). Combining survey data and administrative register data in a multilevel dataset, we are able to control very robustly for the specific characteristics of the students (n = 5,631), the schools (n = 85), and other teacher characteristics (n = 694) besides PSM. We find that PSM is positively associated with examination marks. The result indicates that PSM may be relevant for performance improvements.

Key words: Public service motivation, performance, examination marks, teachers

As the provision of public service is community oriented in nature, "an individual's orientation to delivering services to people with the purpose of doing good for others and society"—also known as public service motivation (PSM)—has attracted considerable interest among public management scholars (Hondeghem and Perry 2009: 6; Perry, Hondeghem and Wise 2010). Indeed, one of the central driving forces behind PSM research has been that PSM is expected improve the performance in public organizations (Perry and Wise 1990). Recent research indicates that managers can actually affect PSM (Jacobsen et al. 2013; Wright et al. 2011), which makes PSM a hidden potential in organizations where goal attainment means doing good for others and/or society. Especially if their specific job also allows them to do good for others and society (Bright 2007; Christensen and Wright 2011; Steijn 2008), public service motivated employees in such organizations are expected to work better and harder and obtain better results. Still, before we can utilize this potential, the causal association between PSM and performance must be firmly established.

Existing studies of the relationship between PSM and performance indicate a positive association, but the heavy reliance on self-reporting and cross-sectional data in these studies renders causal inference difficult (Brewer 2008; Leisink and Steijn 2009; Naff and Crum 1999; Perry, Hondeghem, and Wise 2010: 685; Petrovsky and Ritz 2010; Vandenabeele 2009). To our knowledge, there is only one PSM study using objective performance indicators: Using data from a field experiment of voluntary contributions made by nurses to emergency aid at their own hospital, Bellé (2013) finds that PSM positively affects job performance in a voluntary context. The contribution of this article is to investigate whether PSM also affects performance in terms of the results of core welfare service provision when they are measured objectively.

To be objective, a measure of performance must involve a precise assessment of a given dimension of performance and an external process to verify its accuracy (Perry, Hondeghem, and Wise 2010: 16). Public service delivery outcomes often result from an aggregate effort where several employees work together, and public organizations often have multiple and unclear goals (Dixit 2002), so finding an objective outcome measure which can be combined with information about the individual motivation of the relevant public employee is like finding a needle in a haystack. Due to the extremely high quality of the Danish administrative registers, however, it is possible to link the motivation of individual Danish school teachers to objective performance indicators at the individual level by combining survey data on each teacher's PSM with administrative data on the final examination marks of their students. The fact that children are taught in the same classes by different teachers in different subjects (Danish, math etc.) means that we can control for student-specific characteristics and teacher selection effects by including student fixed effects in the analyses. Controlling for student- and class-specific variation ensures that the estimates of the associations are not biased by school, student, or class level confounding. Due to our research design, we can contribute to answering one of the most central questions in PSM research: Are PSM and individual performance positively related, also when performance is objectively measured?

Our key argument is that PSM makes the individual try harder to do "good for others and society." Especially for the direct service producers in organizations delivering core public services, high PSM means that there is a person–environment fit (i.e., both person–organization and person–job). Provided that individuals see what the organization rates as high performance as being desirable for society and others, we expect PSM to be positively associated with performance. In the following, we first discuss the state of the art in the literature on PSM and individual performance and formulate specific hypotheses followed by a presentation of the data and methods. We then present and discuss the results, and the article concludes with a discussion of its contributions, limitations, and implications.

Public service motivation and individual performance: Theory and state of the art

Perry and Wise (1990) already hypothesized in 1990 that PSM is positively related to individual performance. They argued that public jobs would be intrinsically motivating for individuals with high PSM, because they would embrace work with attributes such as high task significance (Perry and Wise: 371; Perry, Hondeghem, and Wise 2010). For the four dimensions seen as constituting PSM (Perry 1996), the logic is as follows: The first dimension, self-sacrifice, represents the basic pro-social origins of PSM and implies a willingness to deliver services without tangible personal rewards (Perry 1996; Kim and Vandenabeele 2010). For teachers, this could be the willingness to make an extra effort on the job, even if this means that they must put aside personal interests such as time with family and friends. Examples of this are teachers who spend extra (spare) time meeting with parents and students or preparing their teaching. The other three PSM dimensions—“compassion,” “attraction to public policy making,” and “commitment to the public interest”—can be seen as based on affective, instrumental/rational, and normative motives, respectively (Perry and Wise 1990; Wise 2000). Affective motives are based on identification and emphasize an individual’s commitment to or concern for the needs of specific individuals and groups. Here, affective bonding is the emotional basis of serving others, and it is this identification which creates a willingness to do good for others (Kim and Vandenabeele 2010). Among teachers, for example, individuals with a high level of compassion will tend to empathize with children in difficult situations and therefore invest more energy into improving their situation.

Instrumental/rational motives are based on an understanding of how means and measures can be combined in order to contribute to the delivery of public services. For teachers, this could be instrumental participation in decision-making processes in order to do good for the students, for example by trying to affect resource allocation. Finally, norm-based motivation concerns conforming to values and social norms regarding appropriate behavior and societal contributions. In the case of teaching, academic qualifications are valued by society and seen as contributing to the common good, and teachers can be willing to acquire a sense of moral satisfaction (aka. the "warm glow of giving," see Kahneman and Knetsch 1992: 64) by increasing their effort to improve the students' qualifications. These dynamics comprise the theoretical underpinnings of why PSM can be expected to be positively related to performance for jobs (e.g., teaching) where "high performance" means doing good for others and society. However, employees' needs, desires, and preferences are to a higher or lower degree met by the jobs they perform (Kristof-Brown, Zimmerman, and Johnson 2005: 306), and individuals with higher PSM are only likely to make greater effort if their jobs allow them to "do good for others and society" (Hondeghem and Perry 2009: 6); that is, if the PSM fit is high. A PSM fit can accordingly be defined as "comparability between the needs of individuals to serve the public interest and the environmental conditions in their organisation which affect the fulfilment of these altruistic motives" (Taylor 2008: 71–2).

Bright (2007) has argued that diverging results regarding the PSM–performance relationship could be due to the fact that PSM is mediated by a P–E fit (Alonso and Lewis 2001; Leisink and Steijn 2009). Person–job and person–organization fits can be measured directly by asking the respondents about their perceptions of the fit or indirectly by keeping the institutional setting fairly constant in the research design, specifying not only the sector, but also keeping the job context constant by investigating employees who produce very similar services (Christensen and Wright 2011). Variation in PSM is therefore also variation in the PSM fit. Empirical studies using either of the two strategies have tended to find a positive PSM–performance relationship.

Petrovsky and Ritz (2010) thus reviewed the empirical research on PSM and subjective indicators of performance and found that the 10 peer-reviewed studies in the field all find positive associations, while none reports negative associations. The subjective performance indicators vary between self-reported individual level performance (Vandenabeele 2009), perceived organizational effectiveness (Brewer and Selden 2000; Kim 2005), and internal efficiency (Ritz 2009). Furthermore, positive associations are found between PSM and self-reported performance ratings by supervisors (Naff and Crum 1999; Camilleri and van der Heijden 2007). In addition to the research based on subjective indicators of performance, studies of PSM and self-reported behavior show that PSM is positively related to whistleblowing (Brewer and Selden 1998), organizational citizenship behavior (Kim 2005; Pandey, Wright and Moynihan 2008), performance information use (Moynihan and Pandey 2010: 859), work effort, and political influence (Frank and Lewis 2004; Pedersen 2013). This body of research indicates that PSM influences behavior, which again can be linked to different dimensions of performance. Self-reported behavior is, however, subject to social desirability bias just like assessments of PSM and performance (Kim and Kim 2012), and this is particularly serious if the dependent and the independent variables are obtained from the same person in the same measurement context using items with similar characteristics (Podsakoff et al. 2003: 885). Common-source bias can potentially generate many false positives (Meier and O'Toole 2012), and the literature has started to handle this by including administrative data on performance. Using information from several health registers, Andersen and Serritzlew (2012) show that high-PSM physiotherapists are more likely to prioritize disabled patients even if it is more demanding (and provides the same fee) to treat these patients than other patients. A cross-sectional design does, however, make it difficult to determine whether behavior affects PSM or the other way around, and this can be addressed in experimental studies. One laboratory experiment has found that information on working in a public or private context, respectively, influences the work effort of the test persons (Brewer and Brewer 2011). Laboratory experiments are, however, far from the real working settings of public employees, selection effects among the test-persons might occur, and (in this study as elsewhere) sector is an inaccurate proxy for PSM (Christensen and Wright 2011). The external validity is therefore low, even if the causal claims are strong. In a randomized field experiment, Bellé (2013) finds positive PSM–performance relationships using multiple and very good objective output measures. This experiment is conducted in a real-life setting wherein nurses in a public hospital provide humanitarian aid, and this increases the external validity, while the selection bias might still occur, as the participants were randomly selected among volunteers rather than chosen from the entire population (Bellé 2013: 150). Furthermore, the setting of packing surgery equipment for humanitarian aid differs from a typical work situation in being a voluntary activity (even though it takes place at the hospital where the nurses normally work). While the Bellé study is a major contribution to our understanding of the PSM–performance relationship, a gap remains concerning performance which is part of the normal job, especially in terms of the effects of PSM on the results of public service delivery (outcome).

The person–environment fit discussion above implies that a positive association between PSM and performance requires that the investigated performance measure corresponds with the individual's understanding of what is desirable for society and others. According to Boyne (2003), different stakeholders rarely agree on the goal. Even when a consensual performance measure exists, testing the effect of PSM on performance is difficult because the outcome of public service provision is affected by an array of factors other than PSM, e.g., user characteristics, pay structure, education level, professionalism, private or public ownership, other motivational variables, and institutional context (Flynn 2007; Perry, Mesch, and Paarlberg 2006; Rainey and Steinbauer 1999). Additionally, because of low efficacy, using outcome measures – e.g., student examination marks – is a very conservative test in studies of individual employee performance (Miller and Whitford 2007). These difficulties are possibly an important explanation of why individual performance has typically been measured using self-reporting.

All research designs have trade-offs (Wright and Grant 2010: 692), and the strength of subjective performance data is that it is easy to collect and applicable and comparable across organizational contexts. However, self-reported measures often leave the definition of performance to the individual employees. This implies that a self-reported performance measure may rest on very different conceptions of what high performance actually is. One solution is to measure performance as the supervisors' performance appraisals and promotions of the employees (Alonso and Lewis 2001), but this introduces a potential supervisor bias (because supervisors may favor high-PSM employees (Wright and Grant 2010: 695)) and swaps employee subjectivity for supervisor subjectivity. In general, the studies on PSM and subjective performance suffer from social desirability bias and common source bias, which are serious flaws known to produce Type I errors (Brewer 2008; Kim and Kim 2012; Meier and O'Toole, 2012; Petrovsky and Ritz 2010). Thus, studies using objective measures of performance are greatly required.

Conversely, experimental research is able to make strong causal claims but weak in terms of external validity due to selection effects and the artificial setting of especially lab experiments. In Bellé's study, the performance measure is objective in the sense that it stems from an external data source, involves evaluation by others, and is based on explicit criteria. This is an important step forward but takes place in the context wherein the PSM–performance relationship is most likely to be found. It investigates the voluntary work of volunteers, and performance is measured as short-term work effort in emergency aid. Here, altruism is extremely visible and tangible compared to core public services such as education.

The strengths of this article are that it uses an objective outcome indicator, handles selection effects in a real-life setting by using (student) fixed effects methods (see below), and extends the study of PSM and objective performance to a new setting which is at the core of welfare services—the production of public education.

Model and expectations

Investigating the association between individual teachers' PSM and their students' academic performance in Danish public schools, performance is operationalized as students' final examination marks in a given subject. For each observation (a given student's grade in a given subject taught by a given teacher whose PSM we measure), we control for the same student's exam marks in other subjects (taught by other teachers with different PSM levels). Individual performance can be divided into in-role and extra-role behavior (Loon, Vandenabeele, and Leisink 2013; Williams and Anderson 1991), and we use a strictly individual measure of in-role performance; that is, the performance resulting from task-specific role requirements of teaching an individual class. Effects of extra-role behavior, e.g., being a good colleague or contributing to the atmosphere in the working environment, are controlled for via student fixed effects, as explained in detail below. In this manner, the design enables us to focus on the effects of the teachers' in-role performance.

Although grades on exams are a standardized and widely accepted measure of outcome (Andersen and Mortensen 2010; Chubb and Moe 1990; O'Toole and Meier 2011), it is important to stress that Danish schools also have other important objectives than the academic qualifications measured in the final exams (such as promoting a "well-rounded development of the individual student" (Ministry of Children and Education 2012). Almost all public organizations do, however, have multiple objectives, and academic achievement is among the most important goals in Danish schools (Law no. 998, issued August 16 2010). In contrast to Bellé's (2012) study mentioned earlier, this is an outcome measure which is part of the employees' real job, and improvements in performance are likely to demand a long-term work effort.

The discussion of PSM fit implies that the nuances in the PSM–performance relationship may depend on the context of the organization, and a description of Danish schools is therefore necessary to form specific expectations. In Denmark, the 98 municipalities are responsible for providing primary and lower secondary education (from kindergarten to grade 9, ages 6–15). About 85% of Danish children attend public schools which are free of charge, and this study focuses on public schools alone. Public schools are financed by municipal taxes, primarily income taxes, but extensive grants and equalization schemes eliminate the greater part of financial inequalities between municipalities. At the end of compulsory schooling (grade 9), there is an examination in 11 different subjects (see the data section below for details). The examination is closely related to the subjects taught in grades 7–9, especially the 9th grade, and students typically take the examination seriously. Some examinations are written, some are oral, and the degree of external validation differs. Oral exams are conducted by the teacher and an external examiner. Most written examinations are graded by the teacher and an external examiner, but in some subjects they are graded by two external examiners (and not by the teacher). External examiners are appointed by the Ministry of Children and Education.

Teaching involves doing good for students and society, and we accordingly argue that there is a strong PSM fit for Danish teachers with high PSM. Given that it is plausible that all teachers see higher exam marks as desirable, PSM can be expected to have a positive effect on examination marks (Hypothesis 1 below). We also expect that the longer a student has been exposed to a given teacher in a given subject, the stronger the effect; and hence, that children achieve better results the longer they have been taught by high-PSM teachers. Hypothesis 2 (below) expects that the PSM effect is strongest if the students have been taught by a high-PSM teacher throughout lower secondary school (grades 7–9).[1]

This leads to two expectations:

Hypothesis 1: *Children taught by teachers with higher PSM achieve higher examination marks.*

Hypothesis 2: *The positive association between teacher PSM and examination marks is stronger for children taught by the same teacher throughout lower secondary school than children who changed teachers in the relevant subject.*

Method

The main methodological challenge when investigating the causal effect of teacher characteristics on student outcomes is the potential bias arising because the distribution of students and teachers on classes is not random. If, e.g., high-quality teachers sort into classes with more able students (in terms of unobservable characteristics), analyses that fail to

[1] Exposure to a given teacher is conceptualized as teacher change between grades 8 and 9, but the results do not change substantially if it is operationalized as at least one teacher change between grades 7 and 9.

address this sorting pattern would produce upward biased estimates of effects of teacher characteristics. Similarly, compensatory assignment (by policy makers or school administrators) of high-quality teachers to classes with many disadvantaged students may result in downward-biased estimates.

We address the problem of non-random sorting in terms of both observable and unobservable characteristics by including student fixed effects in the empirical models, which is possible because each student has exam marks in multiple subjects and different teachers in different subjects. In our application, the student fixed effects control for all of the unobservable student characteristics (e.g., overall ability and motivation) which are constant across subjects. Only variation between subjects for each student is used in the estimations. The PSM effect on student examination marks is estimated by comparing relative marks in different subjects for each student with the relative PSM of the student's teachers in these subjects (controlling for other variables). Thus, no between-students variation is used. This accounts for the most important part of non-random sorting of students and teachers across classes. However, omitted variable bias may still be a problem, because we cannot control for unobserved student characteristics which are not constant across subjects. If, e.g., high-quality (or high-PSM) teachers are systematically allocated to classes with students having high (or low) ability in the specific subjects taught by these teachers relative to the students' ability in other subjects, our estimates of the effects of teacher characteristics (including PSM) would be biased. Thus, student fixed effects control for potential selection of, e.g., high-ability/high-PSM teachers into classes with overall high-ability students, but not for more specific selection mechanisms where high-ability/high-PSM teachers are selected into classes with high ability in the specific subjects taught by these teachers relative to other subjects. Such subject specific selection mechanisms are very unlikely according to feedback from meetings with a total of more than 800 teachers and principals with whom we have discussed the analysis and results. More importantly, our method does not control for unobserved teacher characteristics, e.g., basic ability to teach. This is a problem if these characteristics are both correlated with PSM and affect student outcomes. However, this basic problem is shared by all other studies we know of which estimate effects of employee motivation or similar characteristics, and we have included control variables for the personal characteristics of the teachers which existing studies indicate are most important: Gender, education, and age or years of teaching experience. These are standard control variables in studies of effects of teacher characteristics (e.g., Bressoux, Kramarz, and Prost 2009; Clotfelter, Ladd and Vigdor 2007a, 2007b, 2010). Age and PSM are expected to be positively correlated due to the increase in generativity (concern for establishing and guiding the next generation) as people get older (Pandey and Stazyk 2008: 102). Females may have higher PSM (Pandey and Stazyk 2008: 102; Perry 1997), but DeHart-Davis, Marlowe, and Pandey (2006) have argued that compassion is a feminine dimension of PSM and that women are no less committed to public service but less likely to declare their commitment to the public interest because of their loyalties to interests in the private realm. Education is relevant because the level of professionalism might affect both PSM and performance (Andersen and Pedersen 2012).

The 'across-subjects' student fixed effects identification strategy is also applied in Clotfelter, Ladd, and Vigdor (2010). Student fixed effects are often used in a different setting wherein longitudinal data on student test scores in a given subject is available for each student for multiple years (e.g., Clotfelter, Ladd, and Vigdor 2007a, 2007b; Rivkin, Hanushek, and Kain 2005). In that case, student fixed effects control for all of the timeinvariant unobservable characteristics of students (such as ability or motivation) that could be correlated with teacher characteristics.

To be specific, we estimate models of the form

$$y_{ijsk} = \alpha + T_{ijsk}\beta + \gamma_s + X_{ijsk}\delta + \mu_i + u_{ijsk} \ , \quad (1)$$

where y_{ijsk} is the examination mark of student i in subject s taught by teacher j in school k, T_{ijsk} is a vector of teacher characteristics, γ_s are subject fixed effects, X_{ijsk} is a vector of interaction terms between student characteristics, teacher characteristics, and subject dummy variables, μ_i are student fixed effects, u_{ijsk} is the error term, α is the constant term, and β and δ are vectors of parameters. Note that even though it is not possible to include the main effects of student characteristics (which do not vary by subject) in the model (because of the student fixed effects) it is possible to include interaction terms between student and teacher characteristics and between student characteristics and subject dummy variables, since these will vary by subject for a given student. All of the estimations include interaction terms between dummy variables for subjects and student gender and immigrant status, since, e.g., boys typically have a comparative advantage in math. The model may be estimated by OLS including a large set of student dummies or by within-student estimation (which produces identical results); see, e.g., Wooldridge (2010).

We have data for teachers in three grades (7–9). In principle, we could include teacher characteristics for each grade as separate variables, but a high correlation would be present in these variables, because many students have the same teacher in a given subject in all three years. In our main analysis, we only use the characteristics of 9th grade teachers but include a dummy variable for change of teachers between grades and interaction terms between this variable and all teacher characteristics. In robustness checks, we have replaced grade 9 teacher characteristics with average characteristics of teachers of grades 7–9, and this gives the same results.

If all subjects are taught in the same basic classes (which is the case in our data), student fixed effects will also take class-specific characteristics into account, including class size and peer group characteristics, provided that the effects of such characteristics do not vary by subject. The institutional feature that different subjects are taught in the same basic classes allows us to identify teacher effects more clearly than in, e.g., Clotfelter, Ladd, and Vigdor (2010). In our analysis, student fixed effects also control for variables at higher levels which do not vary between subjects (e.g., school and municipality fixed effects). This is the case because each student included in our analysis can be registered at only one school.

We estimate robust standard errors taking clustering in schools into account. This is a very aggregate level of clustering and produces conservative standard errors. It takes the inherent two-way clustering in our analysis into account: clustering on teachers (since each teacher teaches many students, in some cases even several classes and/or subjects) and clustering on students (since marks in different subjects for the same student are correlated). Clustering on schools produces robust standard errors corresponding to using a model with random effects at the student, teacher, classroom, and school levels; see, e.g., Kane, Rockoff, and Staiger (2008).

Data

The investigation is based on three sources of data: First, a survey of all teachers from 85 schools conducted from December 2010 until June 2011. Second, to link each individual teacher with each individual student in each subject, we obtained information from the 85 schools on the distribution of students and teachers on classes in grades 7–9 for the three cohorts of students completing 9th grade in 2009, 2010, and 2011. Different subjects are typically taught by different teachers, and we identified all of the teachers who taught the selected cohorts of students in nine subjects which they are examined in: Danish, math, English, history, science, biology, geography, religion, and social studies. In this dataset, teachers are identified by names and initials, providing a link to the teacher questionnaires, which also contain this information. Each student is identified using a unique personal identification number, which enables us to link to administrative register information on students—the third source of data. The register data contains each student's 9th grade examination marks (our performance measure), personal characteristics (e.g., gender, age, ethnic background), and socioeconomic variables for parents (e.g., education, income, family structure, labour market status, working experience).

The survey data from teachers was collected at school staff meetings, where the school principal agreed to let the teachers answer the questionnaire at the meeting. Teachers who were absent from the meetings received a questionnaire and a return envelope. The data from the questionnaires and schools (about the links between students and teachers) were collected by 10 research assistants at the Danish Institute for Government Research and Aarhus University. Student information with personal ID numbers allowed us to link to the administrative register data (with permission from the Danish Data Protection Agency).

The teacher survey and data collection at schools were conducted at relatively large schools to maximize the number of teachers investigated for the given amount of resources available for the data collection. The 221 largest schools in Denmark were contacted, and 38% (85 schools from around Denmark) participated. Most of the remaining 62% did not participate because they did not have a staff meeting within the timeframe (December 2010 to June 2011) with enough time for teachers to complete the questionnaire. Using the administrative register data which is available for all schools, we have tested whether the student characteristics and average exam marks for 9th grade students are different for the 85 participating schools compared to the non-participating schools. The differences are small, but some are significant according to two-sample t tests; especially average exam marks are slightly higher for participating schools (about 0.04 standard deviations in the distribution of individual marks), and the share of students whose parents are immigrants is slightly lower (4.7% compared to 6.0% for nonparticipants). The 85 schools are not representative of all Danish schools because they are bigger. However, if the effects of PSM are different at large schools, the fact that we have selected large schools is an advantage in terms of generalizability to other countries, given that Danish schools are generally smaller than schools in most other countries (Little 2008).

The response rate among staff meeting participants at each school is very close to 100% (only a couple of teachers would not answer). After a review of the data quality, where suspicious entries were deleted, 3,230 usable responses were retained. 1,383 of these teachers had students who were taking their final exams in one of the three investigated years, grades 7–9, and taught them in one of the nine subjects upon which we focus; 1,188 of these teachers had students from these cohorts in grade 9. For the 2011 cohort analysed in this paper, there were 766 9th grade teachers. In the online Appendix, a detailed description of the connection between the three sources of data and the number of observations can be found together with exact wording of the survey questions regarding PSM and the measurement statistics for these. The questions were based on prior surveys (Andersen and Pedersen 2012; Perry 1996), and the final questionnaire was adjusted after a pilot survey of 61 teachers in two schools. The scores on the four PSM dimensions were calculated in a confirmatory factor analysis, controlled for schools. RMSEA is 0.027 (90% C.I. 0.021-0.032), CFI is 0.954, TLI is 0.944, and SRMR indicator is 0.04, indicating that the model has a good fit. PSM is an unweighted sum index of the four dimensions.

The dependent variables in our analyses are examination marks in nine subjects: Danish, math, English, history, science, biology, geography, religion, and social studies. The marks in Danish used here are an average of four individual marks (in reading, spelling, essay, and oral) and math marks are an average of two individual written marks. The exams in Danish and math and the oral examinations in science and English are mandatory. All students must take two additional exams, a written exam in biology or geography, and a written exam in English or an oral exam in history, religion, or social studies. The Ministry of Children and Education decides the distribution of these exams on classes. For students taking the written English exam, the English mark used in the analysis is an average of the marks for the (mandatory) oral exam and the written exam. The reason we use average marks in the case of several individual marks in the same subject (which is the case for Danish, math, and, for some students, English) is that we focus on teacher effects, and teacher characteristics are of course constant within subjects for a given class. Marks are given according to a 7-point scale, but to make interpretation of results easier, we use standardized marks which have mean zero and standard deviation unity for each individual mark in a given subject in a given year (calculated for all students at the 85 schools, not just the estimation sample, which is restricted to observations with teacher information). The descriptive statistics for the dependent variables (examination marks) are shown in the top rows of Table 1, which also lists the explanatory variables used in the analyses.

[Table 1 here]

We did not obtain data for teacher characteristics before grade 7. In principle, this may be a weakness, since these earlier teacher characteristics possibly also affect skills at the end of grade 9. A standard way of handling this kind of problem is to estimate the effect of school inputs (in this case, teacher characteristics) on achievement gains from an earlier to a later grade (the value-added approach; e.g., Todd and Wolpin 2003). We are not able to use this strategy since we have no information on student academic achievement before the examination at the end of grade 9. In practice, it is not an important weakness because of our identification strategy (student fixed effects, see above), and because the more basic skills taught in earlier grades are not very closely related to the examination at the end of grade 9.

Results

Table 2 shows the main estimation results for the effects of the PSM variable on examination marks using student fixed effects models. The estimations in models 1 and 3 are for all exam marks, while the estimations in models 2 and 4 are for written exam marks only. The key variables are PSM and the interaction term between PSM and the "change of teachers" variable. "Change of teachers" is operationalized as a dummy for whether there was a change of teachers between grades 8 and 9. The interaction between PSM and "change of teachers" is included in models 3 and 4 together with interaction terms between "change of teachers" and the other teacher characteristics (last-mentioned not shown). All the models control for teacher characteristics (gender, education, experience), subjects and interaction terms between subjects, and student gender and immigrant status (not shown). A table with all of the coefficients can be accessed in the online Appendix, which also contains robustness checks and tests for non-linearity of the PSM effect.[2]

[Table 2 here]

Hypothesis 1 expects that children taught by teachers with higher PSM get higher examination marks. Model 1 in Table 2 shows a clearly significant and positive association between the PSM of 9th grade teachers and their students' examination marks, and this does not change in model 3, where the interaction between PSM and change of teachers is included. The point estimate for students without a teacher change is 0.034 (see model 3, Table 2). Thus, the effect size is about 0.04, since the standard deviation of the PSM index is 1.26, and that of examination marks is about 1 (see Table 1). The estimated effect of PSM on the written exam marks (models 2 and 4 in Table 2) is slightly smaller (not significantly) than the corresponding estimates for all exams, and it is only significantly different from zero in model 4, where the interaction term between PSM and change of teachers accounts for the fact that not all students have been exposed to the same teacher in grades 7–9.

Hypothesis 2 expects the positive association between teacher PSM and examination marks to be stronger for children taught by the same teacher throughout lower secondary school than children who experienced teacher change. As expected, the point estimate for the interaction term between PSM and "change of teacher" is negative both for all exams (model 3) and for written exams (model 4), lending some support to Hypothesis 2, although the interaction is not statistically significant. For written exams, the interaction term is larger numerically and marginally significant (at the 10% level).

[2] Robustness tests show that results are essentially unchanged if we replace the dummy variable for change of teachers between grades 8 and 9 with a dummy for whether there were at least two different teachers in grades 7–9, or if we replace characteristics of 9th grade teachers with average characteristics of 7th–9th grade teachers. Estimating polynomials of PSM or replacing the PSM index by dummy variables, we do not reject linearity restrictions. Estimating models with interaction effects between PSM and dummies for individual subjects, we cannot reject a hypothesis that all these interaction effects are zero, which indicates that PSM effects are approximately constant across subjects.

Not many variables for 9th grade teacher characteristics other than PSM are significant. The point estimates for the female teacher variable and its interaction with female students indicate that having a female teacher is an advantage for female students but a disadvantage for males. Less than five years of experience affects exam marks negatively, whereas there is no significant difference between having 5–9 years of experience and more than 9 years; these results are consistent with earlier findings, e.g., Clotfelter, Ladd, and Vigdor (2007b).

The number of (student-by-subject) observations and the number of different students, classes, and teachers in the estimation sample are shown at the bottom of Table 2 (more detail is available in the online Appendix). The number of student-by-subject observations is reduced considerably (by 44%) when the analysis is restricted to written marks since marks in many subjects are based on oral exams, but the number of different students and classes is reduced by only 4% and the number of different teachers by 27%. Table 2 presents both the adjusted R^2 of the OLS regressions with dummies for each student and the "R^2 within students" of the regressions on the within-students transformed variables (producing exactly the same parameter estimates). It is hardly surprising that the last goodness of fit measure is much smaller, since R^2 within students is a measure of the fraction of the variation in marks between subjects for individual students which the model can explain. F tests strongly reject the hypothesis that all 5,630 (or 5,421) student fixed effects are zero ($p < 0.0001$). In sum, both main analyses and robustness tests clearly supported Hypothesis 1 and give some (but not decisive) support to Hypothesis 2. Children taught by teachers with higher PSM get higher examination marks, and this association tends to be stronger the longer the children have been taught by teachers with higher PSM.

Discussion of causality: Does PSM *affect* performance?

Our specific research question is whether PSM and individual performance are positively related, also when performance is objectively measured, and the result section above accordingly discusses associations rather than causal effects. But the most interesting question, which appears in the title, is whether teacher PSM (causally) affects student performance in schools (and in other organizations). Given that we use a cross-sectional survey design, we cannot draw firm conclusions regarding causality due to potential endogeneity problems. We will, however, argue below that the two broad classes of rival explanations (reverse causality and omitted variables) are less problematic for this study than for many existing studies of PSM and performance.

The first rival explanation is that high performance may strengthen PSM, whereas low performance weakens PSM instead of the proposed effect of PSM on performance (Wright and Grant 2010: 695). We do not have panel data with information about the same teachers' PSM over time, but we do have data on their students' examination marks for three years, 2009–11. For 2011, PSM is measured before the examination marks are given, and this is the analysis presented in this paper. If the PSM–examination marks correlation is due to reverse causality, we would expect a higher correlation between PSM and the marks of earlier cohorts of students (at least if the mechanism was that marks were seen as a signal/performance information, because then the stochastic variation in the marks in the earlier years would be positively associated with PSM while stochastic variation in the 2011 marks would be uncorrelated with PSM). When rerunning the analyses using data for all three years, however, the estimated PSM coefficients are smaller (approximately 30% smaller) compared to the estimations with only 2011 data (while the t-values are approximately the same, since the standard errors are smaller in the analyses with more observations). This indicates that reverse causality is not a problem. Note that we do find a (smaller) positive correlation between PSM measured in 2011 and performance (student marks) measured in earlier years, but this would also be expected without reverse causality, because PSM is probably rather persistent from year to year.

The second rival explanation is that there is a common cause for both PSM and performance. It can potentially be caused by the non-random selection of students and teachers into classes; as argued above, however, our student fixed effects approach effectively controls for this. However, we cannot rule out that unobserved teacher characteristics might be correlated with both PSM and performance. Wright and Grant (2010: 695) thus argue that conscientiousness (together with supervisor biases, which are discussed below) is a very important omitted variable in many studies. The personality trait "conscientiousness" refers to the degree to which individuals tend to be industrious, disciplined, goal oriented, and organized. Given that it is a robust predictor of job performance across a wide range of occupations and given that there is reason to believe that conscientiousness will be positively associated with PSM, because a sense of duty and responsibility to others is one of the defining features of conscientiousness, Wright and Grant argue (2010: 695) that researchers should examine whether PSM predicts higher performance even after controlling for conscientiousness. A counter argument is that it would be difficult and not necessarily fruitful due to the conceptual overlap between the concepts.

The fact that we measure performance objectively, using administrative register data, also gives us more reason to believe that the association can be causal. Avoiding social desirability and supervisor biases in the dependent variable ensures that we do not have these types of omitted variable bias. Many of the studies linking PSM to performance have measured performance as the supervisors' performance appraisals and/or employee promotions (e.g., Alonso and Lewis 2001), and Wright and Grant (2010: 695) argue that an alternative explanation for the identified positive association (in addition to confounding from conscientiousness) is that supervisors are biased in favor of high-PSM employees and award them in performance appraisals by giving them more credit for their contributions. As Wright and Grant (2010) argue, this may skew objective performance measures if supervisors offer employees with high PSM more resources and support than low-PSM employees. This is not a serious problem in Danish schools, because rigid workload agreements between the teachers' union and the municipalities regulate the time allotted for teaching, and resources available for teaching vary little among the teachers at the same school. We agree with Wright and Grant (2010: 696) that randomized, controlled field experiments with interventions designed to increase PSM would be very desirable in further studies of the PSM–performance association, but we also argue that this study brings us a giant step forward by conducting a test of the PSM–performance relationship measured in a very objective manner. Although we cannot say for sure whether PSM causally affects performance in schools, our results using this research design with student fixed effects and objective performance data are at least consistent with a positive causal effect.

Conclusion

We set out to investigate whether students taught by teachers with higher PSM receive higher examination marks (i.e., perform better), and our key finding is clear: In a context in which performance can obviously be seen as benefitting others and society, higher PSM is associated with higher performance, and this result is stronger the longer the students have been taught by a high-PSM teacher. In relation to the existing literature, this contributes in at least three ways: First, our results strongly indicate that PSM affects the outcome of ordinary, everyday public service provision. This adds to a literature which has demonstrated that PSM affects behavior (Andersen and Serritzlew 2012; Pandey, Wright and Moynihan 2008), self-reported performance (Brewer 2008; Leisink and Steijn 2009; Vandenabeele 2009), and objectively measured performance in a voluntary context (Bellé 2012). Second, the article draws attention to the importance of the length of time to which users of public services have been exposed to employees with PSM. Many important public services, especially education, take a long time to produce, and this study indicates that PSM is also about the long haul, suggesting that studies investigating the effect of PSM over time would be extremely useful. Third, the article demonstrates a strategy for handling the person–environment question (that PSM can only be expected to affect performance if there is a PSM fit; Bright 2007; Christensen and Wright 2011; Taylor 2008) by keeping the context approximately constant for a public service where it can hardly be questioned whether high performance (student performance in the final exams) is equivalent to doing good for others and society.

Our test is strong in several ways. First, the method (student fixed effect regression) very robustly ensures that student background and selection effects do not confound the results. Second, and most importantly, we show that the positive PSM–performance association identified in studies using self-reported performance measures is also found in this study, in which the performance measure is objective and based on register data. We thus solidify the link between PSM and performance. Still, our study has limitations. Part of the estimated "effect" of PSM on the objective performance measure could be due to unobserved teacher characteristics which are correlated with both PSM and performance. The study holds the institutional context constant (at the school, municipal, and country levels), which strengthens the internal validity; but in future studies it would be interesting to analyze the relationships between context, PSM, and performance. In terms of external validity, the specific results cannot be statistically generalized outside Denmark, but nothing indicates that the overall findings do not apply in other contexts. Applying the same approach to data in other countries would be interesting, and conducting a similar analysis for other occupations where the individual performance of different public employees can be compared for the same user would increase generalizability. Combined with other studies of the PSM–performance relationship, our findings strongly support that PSM is relevant for performance in organizations where the goal is linked to how the employees see “the public good.”

Further perspectives

To fully exploit the potential of PSM, future research should both establish the causal relationship between PSM and performance more firmly and continue to investigate how public service-motivated employees can be attracted, selected, and retained, and how leadership can increase PSM. Furthermore, it would also be useful to analyze directly how the person–job and person–organization fits moderate the PSM–performance relationship, as this would enable us to determine in which organizations there is a potential to increase performance through increased PSM. The empirical results show that PSM is related to objective measures of performance. If research finds that the causal relationship is solid, however, the practical implications for public administration are substantial.

First, the public sector may have a hidden potential if PSM is consciously used in the attraction, selection, and attrition of public sector employees (Perry and Wise 1990).

In 2010, Wright and Grant (2010: 693) argued that we knew too little about the stability of PSM to use PSM in managerial decisions, but recent research has answered their call for more research, finding that PSM seems to be a dynamic state rather than a stable trait or a disposition (Kjeldsen and Jacobsen 2012; Bellé 2012). This means that there are several managerial and administrative implications of PSM–performance relationship. Managers can increase the level of PSM among their employees through the mechanisms of attraction, selection, and attrition; for instance, they can design job packages which are attractive to high-PSM employees (Andersen et al. 2012). Second, managers can do more to avoid employee PSM being crowded out by incentives and command systems. Recent research indicates that they can avoid this by implementing these systems in a way that makes employees perceive them as supportive rather than controlling (Andersen, Kristensen, and Pedersen 2011; Jacobsen, Hvitved, and Andersen 2013). The existence of a causal relationship between PSM and performance makes the potential gains of this type of management more evident. Third, managers in organizations without severe value conflicts can use transformational leadership to increase employee PSM (and thus increase performance), given that existing studies reveal a positive association between transformational leadership and PSM for these organizations (Krogsgaard, Thomsen, and Andersen 2013; Park and Rainey 2008; Wright, Moynihan, and Pandey 2011).

References

Alonso, Pablo, and Gregory B. Lewis. 2001. Public service motivation and job performance: Evidence from the federal sector. *American Review of Public Administration* 31 (4):363–80.

Andersen, Lotte Bøgh, Tor Eriksson, Nicolai Kristensen, and Lene Holm Pedersen. 2012. Attracting public service motivated employees: How to design compensation packages in diverse societies. *International Review of Administrative Sciences* 78:615–41.

Andersen, Lotte Bøgh, Nicolai Kristensen, and Lene Holm Pedersen. 2011. *Do documentation requirements reduce intrinsic motivation and increase worker absence?* 11th Public Management Research Association Conference, Syracuse, June 2–4, 2011.

Andersen, Lotte Bøgh, and Lene Holm Pedersen. 2012. Public service motivation and professionalism. *International Journal of Public Administration* 35 (1):46–57.

Andersen, Lotte Bøgh, and Søren Serritzlew. 2012. Does public service motivation affect the behavior of professionals? *International Journal of Public Administration* 35 (1):19–29.

Andersen, Simon Calmar, and Peter B. Mortensen. 2010. Policy stability and organizational performance: Is there a relationship? *Journal of Public Administration Research & Theory* 20 (1):1–22.

Bellé, Nicola (2013): Experimental Evidence on the Relationship between Public Service Motivation and Job Performance. *Public Administration Review*, 73(1) 143-153

Boyne, George A. 2003. What is public service improvement? *Public Administration* 81(2):211–27.

Bressoux, Pascal, Francis Kramarz, and Corinne Prost. 2009. Teachers' training, class size and students' outcomes: Learning from administrative forecasting mistakes. *The Economic Journal* 119:540–61.

Brewer, Gene A. and Sally C. Selden (1998): Whistle-blowers in the Federal Civil Service: New Evidence of the Public Service Ethic. *Journal of Public Administration Research and Theory* 8:413-39.

Brewer, Gene A. 2008. Employee and organizational performance. In *Motivation in Public Management: The Call of Public Service*, eds. James L. Perry, and Annie Hondeghem, 136–56. Oxford: Oxford University Press.

Brewer, Gene A., and Gene A. Brewer Jr. 2011. Parsing public/private differences in work motivation and performance: An experimental study. *Journal of Public Administration Research & Theory* 21:347–62.

Brewer, Gene A., and Sally Coleman Selden. 2000. Why elephants gallop: Assessing and predicting organizational performance in federal agencies. *Journal of Public Administration Research & Theory* 4:685–711.

Bright, Leonard. 2007. Does person–organization fit mediate the relationship between public service motivation and the job-performance of public employees? *Review of Public Personnel Administration* 27:361–79.

Christensen, Robert K., and Bradley E. Wright. 2011. The effects of public service motivation on job choice decisions: Disentangling the contributions of person-organization fit and person–job fit. *Journal of Public Administration Research & Theory* 21:723–43.

Camilleri, E., and B. I. J. M. van der Heijden. 2007. Organizational commitment, public service motivation, and performance within the public sector. *Public Performance & Management Review*, 31(2):241–74.

Chubb, J. E., and T. M. Moe. 1990. *Politics, Markets, and America's Schools*. Washington, DC: Brookings Institution.

Clotfelter, Charles T., Helen F. Ladd, and Jacob L. Vigdor. 2007a. How and why do teacher credentials matter for student achievement? *NBER Working Paper 12828*.

———. 2007b. Teacher credentials and student achievement: Longitudinal analysis with student fixed effects. *Economics of Education Review* 26:673–82.

———. 2010. Teacher credentials and student achievement in high school: A cross-subject analysis with student fixed effects. *Journal of Human Resources* 45(3):655–81.

DeHart-Davis, Leisha, Justin Marlowe, and Sanjay K. Pandey. 2006. Gender dimensions of public service motivation. *Public Administration Review* 66(6):873–87.

Dixit, Avinash. 2002. Incentives and organizations in the public sector: An interpretative review. *Journal of Human Resources* 37:696–727.

Flynn, Norman. 2007. *Public Sector Management*, 5th ed. London: Thousand Oaks, New Delhi: Sage.

Frank, Sue A., and Gregory B. Lewis. 2004. Government employees: Working hard or hardly working? *The American Review of Public Administration* 34:36.

Hondeghem, Annie, and James L. Perry. 2009. EGPA symposium on public service motivation and performance: Introduction. *International Review of Administrative Sciences* 75:5–9.

Jacobsen, Christian Bøtcher, Johan Hvitved, and Lotte Bøgh Andersen. 2013. Command and motivation: How the perception of external interventions relates to intrinsic motivation and public service motivation. *Public Administration.* Published online 25 March 2013.

Kahneman, Daniel, and Jack L. Knetsch. 1992. Valuing public goods: The purchase of moral satisfaction. *Journal of Environmental Economics and Management* 22:57–70.

Kane, Thomas J., Jonah E. Rockoff, and Douglas O. Staiger. 2008. What does certification tell us about teacher effectiveness? Evidence from New York City. *Economics of Education Review* 27:615–31.

Kim, Sangmook. 2005. Individual-level factors and organizational performance in government organizations. *Journal of Public Administration Research & Theory* 15(2):245–61.

Kim, Seung Hyun, and Sangmook Kim. 2012. *Cultural diversity and social desirability bias in measuring public service motivation*. EGPA Conference, Bergen, Norway, September 5–8, 2012.

Kim, Sangmook, and Wouter Vandenabeele. 2010. A strategy for building public service motivation research internationally. *Public Administration Review* 70:701–9.

Kjeldsen, Anne Mette, and Christian Bøtcher Jacobsen. 2012. Public service motivation and employment sector: Attraction or socialization? *Journal of Public Administration Research & Theory.* Published online October 29 2012.

Kristof-Brown, Amy L., Ryan D. Zimmerman, and Erin C. Johnson. 2005. Consequences of individuals' fit at work: A meta-analysis of person–job, person–organization, person–group, and person–supervisor fit. *Personnel Psychology* 58(2):281–342.

Krogsgaard, Julie Alsøe, Pernille Thomsen, and Lotte Bøgh Andersen. 2013. *Only if we agree? How value conflict moderates the relationship between transformational leadership and public service motivation*. IRSPM conference, Prague 10–12 April 2013.

Leisink, Peter, and Bram Steijn. 2009. Public service motivation and job performance of public sector employees in the Netherlands. *International Review of Administrative Sciences* 75:35–52.

Little, Angela W. 2008. Size matters for EFA. *Research Monograph No 26*. http://www.create-rpc.org/pdf_documents/PTA26.pdf (accessed May 23 2012).

Loon, Nina van; Wouter Vandenabeele, and Peter Leisink. 2013. *PSM's effect on performance: What do we see when including fit and distinguishing different dimensions of performance?* IRSPM conference, Prague 10–12 April 2013.

Meier, Kenneth J., and Laurence J. O'Toole. 2012. Subjective organizational performance and measurement error: Common source bias and spurious relationships. *Journal of Public Administration Research & Theory*. First published online December 8, 2012.

Miller, Gary J., and Andrew B. Whitford. 2007. The principal's moral hazard: Constraints on the use of incentives in hierarchy. *Journal of Public Administration Research and Theory* 17(2):213–33.

Ministry of Children and Education. 2012. *The Aims of the "Folkeskole"* http://eng.uvm.dk/Education/Primary-and-Lower-Secondary-Education/The-Folkeskole/The-Aims-of-the-Folkeskole (accessed October 9 2012).

Moynihan, Donald P., and Sanjay K. Pandey. 2010. The big question for performance management: Why do managers use performance information? *Journal of Public Administration Research and Theory* 20(4):849–66.

Naff, Katherine C., and John Crum. 1999. Working for America: Does public service motivation make a difference? *Review of Public Personnel Administration* 19(4):5–16.

O'Toole, Laurence, and Kenneth Meier. 2011. *Public Management: Organizations, Governance, and Performance.* Cambridge: Cambridge University Press.

Pandey, Sanjay K., and Edmund C. Stazyk. 2008. Antecedents and correlates of public service motivation. In *Motivation in Public Management: The Call of Public Service*, eds. James L. Perry and Annie Hondeghem, 101–17. Oxford: Oxford University Press.

Pandey, Sanjay K., Bradley E. Wright, and Donald P. Moynihan. 2008. Public service motivation and interorganizational citizenship behavior: Testing a preliminary model. *International Public Management Journal* 11(1):89–108.

Park, Sung Min, and Hal G. Rainey, Hal G. 2008. Leadership and public service motivation in US federal agencies. *International Public Management Journal* 11(1):109–42.

Pedersen, Lene Holm. 2013. Committed to the public interest? Motivation and behavioural outcomes among local councillors. *Public Administration*, Early view.

Perry, James L. 1996. Measuring public service motivation: An assessment of construct reliability and validity. *Journal of Public Administration Research and Theory* 6(1):5–24.

——— 1997. Antecedents of public service motivation. *Journal of Public Administration Research and Theory* 7(2):181–97.

Perry, James L., Annie Hondeghem, and Lois R. Wise. 2010. Revisiting the motivational bases of public service: Twenty years of research and an agenda for the future. *Public Administration Review* 70: 681–90. doi: 10.1111/j.1540-6210.2010.02196.x

Perry, James L., Debra Mesch, and Laurie Paarlberg. 2006. Motivating employees in a New Governance era: The performance paradigm revisited. *Public Administration Review* 66(4):505–14.

Perry, James L., and Lois R. Wise. 1990. The motivational bases of public service. *Public Administration Review* 50:367–73.

Petrovsky, Nicolai, and Adrian Ritz. 2010. *Do motivated elephants gallop faster? An assessment of the effect of public service motivation on government performance at the individual and organizational levels.* Paper presented to the International Conference on Public Service Motivation (PSM) and Public Performance in a Globalized World, Huazhong University of Science & Technology, Wuhan China, July 15–17, 2010

Podsakoff, Philip M., Scott B. MacKenzie, Jeong-Yeon Lee, and Nathan P. Podsakoff. 2003. Common method biases in behavioral research: A critical review of the literature and recommended remedies. *Journal of Applied Psychology* 88(5):879–903.

Rainey, Hal G., and Paula Steinbauer. 1999. Galloping elephants: Developing elements of a theory of effective government organizations. *Journal of Public Administration Research* 9(1):1–32.

Ritz, Adrian. 2009. Public service motivation and organizational performance in Swiss federal government. *International Review of Administrative Sciences* 75(1):53–78.

Rivkin, Steven, Eric A. Hanushek, and John F. Kain. 2005. Teachers, schools, and academic achievement. *Econometrica* 73(2):417–57.

Steijn, Bram. 2008. Person–environment fit and public service motivation. *International Public Management Journal* 11:13–27.

Taylor, Jeanette. 2008. Organizational influences, public service motivation and work outcomes: An Australian study. *International Journal of Public Management* 11(1):67–88.

Todd, Petra E., and Kenneth I. Wolpin. 2003. On the specification and estimation of the production function for cognitive achievement. *Economic Journal* 113:F3–F33.

Vandenabeele, Wouter. 2009. The mediating effect of job satisfaction and organizational commitment on self-reported performance: more robust evidence of the PSM–performance relationship. *International Review of Administrative Sciences* 75:11–34.

Williams, Larry J., and Anderson, Stella E. 1991. Job satisfaction and organizational commitment as predictors of organizational citizenship and in-role behaviors. *Journal of Management* 17(3):601–17.

Wise, Lois R. 2000. The public service culture. In *Public Administration Concepts and Cases*, 7th ed., ed. Richard J. Stillman II, 342–53. Boston: Houghton Mifflin.

Wooldridge, Jeffrey M. 2010. *Econometric analysis of cross section and panel data*. Cambridge, MA: MIT Press.

Wright, Bradley E., and Adam M. Grant. 2010. Unanswered questions about public service motivation: Designing research to address key issues of emergence and effects. *Public Administration Review* 70(5):691–700.

Wright, Bradley E., Donald P. Moynihan, and Sanjay K. Pandey. 2011. Pulling the levers: Transformational leadership, public service motivation, and mission valence. *Public Administration Review* 72(2):206–15.

Table 1. Summary statistics: student-by-subject observations for the 2011 cohort estimation sample

	count	mean	sd	min	max
All exam marks	24360	0.000	0.958	–3.149	1.955
Written exam marks	13660	0.004	0.942	–3.149	1.955
Teacher characteristics:					
PSM	24360	0.032	1.261	–4.650	3.089
Compassion	24360	–0.018	0.520	–2.683	0.649
Commitment to the public interest	24360	0.005	0.333	–1.435	0.594
Attraction to public policy making	24360	0.020	0.489	–1.007	1.111
Self-sacrifice	24360	0.025	0.504	–1.459	1.197
Female teacher	24360	0.511	0.500	0.000	1.000
Female teacher & student	24360	0.253	0.435	0.000	1.000
Qualifications in subject	24360	0.689	0.463	0.000	1.000
Special teacher education	24360	0.079	0.269	0.000	1.000
No teacher education	24360	0.020	0.140	0.000	1.000
Experience 0–4 years	24360	0.178	0.383	0.000	1.000
Experience 5–9 years	24360	0.260	0.439	0.000	1.000
Change of teachers (from 8^{th} to 9^{th} grade)	24360	0.219	0.414	0.000	1.000
2+ teachers in grades 7–9 in the subject	24360	0.387	0.487	0.000	1.000
3 teachers in grades 7–9 in the subject	24360	0.059	0.235	0.000	1.000
Dummy variables for subjects:					
Danish	24360	0.184	0.388	0.000	1.000
Math	24360	0.182	0.386	0.000	1.000
English	24360	0.186	0.389	0.000	1.000
History	24360	0.040	0.196	0.000	1.000
Science	24360	0.166	0.372	0.000	1.000
Biology	24360	0.078	0.269	0.000	1.000
Geography	24360	0.085	0.278	0.000	1.000
Religion	24360	0.042	0.200	0.000	1.000
Social studies	24360	0.037	0.188	0.000	1.000

Reference categories are: Teacher experience at least 10 years; teacher education standard.

Table 2. Effects of 9th grade teacher PSM on student examination marks: Student fixed effects regressions

	Without interaction between PSM and teacher change		With interaction between PSM and teacher change	
	(1) Exam marks	(2) Written marks	(3) Exam marks	(4) Written marks
PSM	0.0318***	0.0170	0.0344**	0.0235*
	(0.00900)	(0.00970)	(0.0104)	(0.00953)
Female teacher	–0.0148	–0.000567	-0.0251	-0.00595
	(0.0265)	(0.0280)	(0.0266)	(0.0310)
Female teacher & student	0.0484	0.0142	0.0685**	0.0320
	(0.0248)	(0.0302)	(0.0245)	(0.0321)
Qualifications in subject	–0.00401	0.00895	0.00375	0.00232
	(0.0186)	(0.0230)	(0.0223)	(0.0253)
Special teacher edu.	0.0382	0.0465	0.0253	0.0391
	(0.0417)	(0.0529)	(0.0546)	(0.0665)
No teacher education	–0.0789	–0.184	–0.0885	–0.158
	(0.0884)	(0.144)	(0.122)	(0.189)
Experience 0–4 years	–0.0612*	–0.0567	–0.0666*	–0.0577
	(0.0256)	(0.0336)	(0.0290)	(0.0331)
Experience 5–9 years	0.00946	0.00155	0.0198	0.0305
	(0.0240)	(0.0274)	(0.0279)	(0.0292)
Change of teachers	–0.0554*	–0.0430	–0.0279	–0.0262
	(0.0241)	(0.0257)	(0.0402)	(0.0526)
PSM * change of teachers			–0.0131	–0.0338
			(0.0147)	(0.0181)
Observations (students*subjects)	24360	13660	24360	13660
Students	5631	5422	5631	5422
Classes	280	269	280	269
Teachers	694	509	694	509
Schools	85	85	85	85
Adj. R^2 (OLS, student dummies)	0.563	0.648	0.564	0.649
R^2 (within student)	0.034	0.065	0.035	0.068

All estimates include controls for subjects and interaction terms between dummy variables for subjects and students' gender and immigrant status; estimates (3) and (4) also control for interaction terms between teacher characteristics and change of teachers; see Appendix Table A.4 for details.
Standard errors in parentheses—robust standard errors clustered on schools.
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Appendix (to be available online):

Numbers of observations in data from schools, teacher survey, and registers

Table A.1 shows the numbers of students, student–subject observations, teachers, and classes for the three cohorts of students who completed 9th grade in 2009–11. The first four columns show the numbers in the dataset collected from the 85 schools, which links teachers to students. For 9th grade, there are 16,858 students, 147,849 student–subject observations, 1,806 teachers, and 833 classes. Since there are nine subjects, the number of student–subject observations for 9th grade should be nine times the number of student observations (i.e., 151,722), but the observed number (147,849) is about 2.5% smaller, which is due to missing or incomplete data for teachers in some subjects from a few schools. The problem with incomplete data from schools is worse for the 7th and 8th grades, but the main reason for the rather low number of student–subject observations for 7th grade in particular is the fact that not all subjects are taught in all three grades; thus, religion and social studies are typically not taught in 7th grade (in accordance with guidelines from the Ministry of Children and Education). The last four columns in Table A.1 show the corresponding numbers when only considering observations with matched teacher data from the survey. Thus, 1,188 (66%) of the 1,806 teachers who taught the 9th grade responded to the questionnaire and could be linked to the students in the dataset. The 34% missing observations for 9th grade teachers are due to absence from the staff meeting at which the questionnaires were handed out (the response rate of these teachers is rather low). Another important reason is that some of the teachers who taught 9th grade in 2009 and 2010 were no longer at the school in 2011 (because of mobility or retirement). Thus, it is not surprising that missing teacher survey observations is a more serious problem for the 7th and 8th grades than grade 9 or that the problem is less serious for 9th grade teachers in 2011 (the year of data collection), where we have survey data for 75% of the teachers (766 out of 1,017); see Table A.2.

[Tables A.1, A.2 and A.3 here]

Table A.3 shows the number of observations (with survey data on teachers) after merging with register data for students' marks and their respective socioeconomic background. The register data contains all students who have marks for the year's work in one of the years 2009–11. All students with exam marks also have marks for the year's work. A comparison of the first column of Table A.3 and the last column of Table A.1 shows that about 4% of the student (and student–subject) observations are lost due to missing register data (consequently, about 1% of classes and matched teachers are lost). The missing register data may be due to students' school mobility in the months before the exam, dropouts, exemption from examinations for special education students, or errors in the registration of personal identification numbers at schools. Similar percentages of observations with missing register data apply if we consider the 2011 cohort separately; see the fourth column of Table A.3 and the last column of Table A.2. A comparison of columns 1 and 2 (and similarly 4 and 5 for the 2011 cohort) in Table A.3 shows that the number of student–subject observations is much smaller for examination marks than for marks for the year's work. This is because all students completing grade 9 will have marks for the year's work in each subject, whereas not all students will take exams in all subjects, and especially because each student typically only has one exam mark in either biology or geography (but not in both subjects), and only one in either history, religion, social studies, or written English (see the data section). The number of student–subject observations is also reduced considerably when only subjects with written exams are considered; see the third and sixth columns of Table A.3.

Full estimation results for the main regressions

Table A.4 contains the full student fixed effects regression results corresponding to columns (3) and (4) of Table 2.

[Table A.4 here]

PSM factor analysis

Tables A.5 and A.6 present the results of the PSM factor analysis.

[Tables A.5 and A.6 here]

Tests for non-linearity of the PSM effect

We tested for non-linearity of the effect of the PSM index in several ways. First, we included a second-order polynomial of PSM in the models of Table 2. In this analysis, we added a constant term to the PSM index to ensure non-negativity (and thereby monotony of the squared term). To be specific, we added the numerical value of the minimum of PSM so that the PSM index varies in the interval from 0 to 7.739 (instead of the interval from –4.650 to 3.089, see Table 1). Of course, in a linear model this transformation does not affect the estimated parameter of PSM, but only the constant term of the regression. Adding PSM-squared as an extra explanatory variable in the four models of Table 2 does not indicate non-linearity in the effect of PSM: The estimated coefficients of the squared term are very small and insignificant with t values between –0.75 and 0.02, and figures of the estimated polynomials are close to straight lines.

As another type of test for non-linearity, we replaced the PSM index with a set of dummy variables. For instance, we included three dummy variables based on the quartiles of PSM: Choosing values of PSM below the first quartile as reference category, the first dummy is 1 if PSM is in the range between the first and second quartiles, the second dummy is 1 if PSM is in the range between the second and third quartiles, and the third dummy is 1 if PSM is larger than the third quartile. Let c_2, c_3 and c_4 denote the coefficients of these three dummies, and let m_1, m_2, m_3 and m_4 denote the means of the PSM index within each interval determined by the quartiles (e.g., m_1 is the mean of PSM for observations with PSM values below the first quartile, and m_4 is the mean of PSM for observations with PSM values above the third quartile). In this case, linearity imposes two linear restrictions on the dummy coefficients:

$$\frac{c_3 - c_2}{m_3 - m_2} = \frac{c_2}{m_2 - m_1}\ , \quad \frac{c_4 - c_3}{m_4 - m_3} = \frac{c_3 - c_2}{m_3 - m_2}$$

Estimating the four models of Table 2 with the PSM index replaced by the three dummies, we cannot reject these two restrictions implied by linearity: The p values of the F test statistic are between 0.39 and 0.92 in these models.

Table A.1. Numbers of students, student–subject observations, teachers, and classes for all three cohorts

	Dataset from schools				With matched teacher information			
Grade	7–9	7	8	9	7–9	7	8	9
Students	16,858	12,952	15,415	16,858	16,797	12,614	15,212	16,778
Student–subjects	364,819	88,041	128,929	147,849	246,553	53,843	86,820	105,890
Teachers	2454	1605	1842	1806	1383	930	1144	1188
Classes	2286	666	787	833	2254	648	777	829

Table A.2. Numbers of students, student–subject observations, teachers, and classes for the 2011 cohort

	Dataset from schools				With matched teacher information			
Grade	7–9	7	8	9	7–9	7	8	9
Students	5895	5061	5539	5895	5895	4951	5450	5895
Student–subjects	133,548	34,318	46,556	52,674	95,142	21,759	32,658	40,725
Teachers	1442	836	984	1017	909	515	660	766
Classes	826	257	282	287	816	251	278	287

Table A.3. Numbers of 9th grade students, student–subject observations, teachers, and classes after merging with register data for students' marks and socioeconomic background

	2009–11 cohorts			2011 cohort		
	Marks for the year's work	Exam marks	Written exam marks	Marks for the year's work	Exam marks	Written exam marks
Students	16,169	15,986	15,379	5679	5631	5422
Student–subjects	101,598	633,77	36,088	39,056	24,360	13,660
Teachers	1178	1129	972	747	694	509
Classes	820	813	784	281	280	269

Table A.4. Effects of 9th grade teacher PSM on student examination marks: Full Student FE regression results corresponding to columns (3) and (4) of Table 2

	(1) Exam marks	(2) Written marks
PSM	0.0344**	0.0235*
	(0.0104)	(0.00953)
Female teacher	–0.0251	–0.00595
	(0.0266)	(0.0310)
Female teacher & student	0.0685**	0.0320
	(0.0245)	(0.0321)
Qualifications in subject	0.00375	0.00232
	(0.0223)	(0.0253)
Special teacher edu.	0.0253	0.0391
	(0.0546)	(0.0665)
No teacher education	–0.0885	–0.158
	(0.122)	(0.189)
Experience 0–4 years	–0.0666*	–0.0577
	(0.0290)	(0.0331)
Experience 5–9 years	0.0198	0.0305
	(0.0279)	(0.0292)
Change of teachers	–0.0279	–0.0262
	(0.0402)	(0.0526)
PSM * change of teachers	–0.0131	–0.0338
	(0.0147)	(0.0181)
Female teacher * change of teachers	0.0468	0.0360
	(0.0445)	(0.0517)
Female teacher&student * change	–0.0959*	–0.0897
	(0.0454)	(0.0511)
Qualifications in subject * change	–0.0368	0.0404
	(0.0460)	(0.0514)
Special teacher edu. * change	0.0584	0.0910
	(0.0708)	(0.0890)
No teacher education * change	0.0318	–0.0557

	(0.123)	(0.224)
Experience 0–4 yrs. * change	0.0293	–0.00574
	(0.0628)	(0.0656)
Experience 5–9 yrs. * change	–0.0463	–0.150**
	(0.0510)	(0.0567)
Math	0.226***	0.200***
	(0.0304)	(0.0321)
English	0.131***	0.266***
	(0.0301)	(0.0545)
History	0.155***	
	(0.0404)	
Science	0.119***	
	(0.0334)	
Biology	0.247***	0.251***
	(0.0436)	(0.0366)
Geography	0.295***	0.285***
	(0.0333)	(0.0329)
Religion	–0.0231	
	(0.0564)	
Social studies	0.122*	
	(0.0523)	
Math, female student	–0.447***	–0.439***
	(0.0224)	(0.0247)
English, female student	–0.306***	–0.455***
	(0.0253)	(0.0570)
History, female student	–0.310***	
	(0.0465)	
Science, female student	–0.262***	
	(0.0328)	
Biology, female student	–0.413***	–0.420***
	(0.0398)	(0.0416)
Geography, female student	–0.587***	–0.597***

	(0.0317)	(0.0354)
Religion, female student	-0.0689	
	(0.0672)	
Social studies, female student	–0.353***	
	(0.0564)	
Math, immigrant	0.00699	0.0588
	(0.0524)	(0.0558)
English, immigrant	0.268***	0.174
	(0.0521)	(0.128)
History, immigrant	0.259*	
	(0.122)	
Science, immigrant	0.108*	
	(0.0488)	
Biology, immigrant	–0.120	–0.103
	(0.0730)	(0.0709)
Geography, immigrant	–0.126	–0.0554
	(0.0804)	(0.0813)
Religion, immigrant	0.276**	
	(0.0955)	
Social studies, immigrant	0.0348	
	(0.103)	
Constant	0.00941	0.00502
	(0.0270)	(0.0330)
Observations	24,360	13,660
R^2 (within student)	0.035	0.068

Reference categories are: Teacher experience at least 10 years; the subject is Danish; teacher education standard.
Standard errors in parentheses – robust standard errors clustered on schools.
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table A.5: Measurement statistics for PSM dimensions. Structural Equation model (clustered by schools)

Dimensions and items (English translation in italics)	SFL	SMC (R^2)
Self-sacrifice		
Det er vigtigere for mig at gøre en forskel i forhold til samfundet end at opnå personlig vinding. *Making a difference in society means more to me than personal achievements.*	0.584	0.341
Jeg mener, at man skal bidrage med mere til samfundet, end man modtager. *I feel people should give more back to society than they get from it.*	0.546	0.298
Jeg er villig til at risikere at skulle tilsidesætte mine personlige behov for samfundets skyld. *I am willing to risk personal loss to help society.*	0.858	0.736
Jeg er klar til at lide afsavn for samfundets skyld. *I am prepared to make sacrifices for the good of society.*	0.834	0.695
Jeg sætter samfundsmæssige forpligtigelser over hensynet til mig selv. *I believe in putting duty before self.*	0.785	0.616
Compassion		
Jeg bliver følelsesmæssigt berørt, når jeg ser mennesker i nød. *It is difficult for me to contain my feelings when I see people in distress.*	0.724	0.524
For mig er hensyntagen til andres velfærd meget vigtig. *For me, considering the welfare of others is one of the most important values.*	0.709	0.502
Jeg bliver meget berørt, når jeg ser andre mennesker blive behandlet uretfærdigt. *I get very upset when I see other people being treated unfairly.*	0.717	0.514
Jeg føler sympati overfor mindre privilegerede mennesker med problemer. *I feel sympathetic to the plight of the underprivileged.*	0.674	0.454
Attraction to public policy making		
Jeg forbinder generelt politik med noget positivt. *I generally associate politics with something positive.*	0.503	0.253
Jeg bryder mig ikke om politiske studehandler. *The give and take of public policy making doesn't appeal to me (reversed).*	0.381	0.146
Jeg har ikke særligt høje tanker om politikere. *I do not care much for politicians (reversed).*	0.841	0.707
Commitment to the public interest		
Det er vigtigt for mig, at offentlige ydelser gavner samfundet som helhed. *It is important for me that public services contribute to the common good.*	0.434	0.189
Jeg så helst, at offentligt ansatte gør det, der er bedst for hele samfundet, selvom det skulle gå ud over mine egne interesser. *I would prefer seeing public officials do what is best for the whole community, even if it harmed my interests.*	0.567	0.321
Det er vigtigt for mig at bidrage til det fælles bedste. *It is important for me to contribute to the common good.*	0.734	0.539
Det er min borgerpligt at gøre noget, der tjener samfundets bedste. *I consider public service my civic duty.*	0.752	0.566

Note: SFL = standardized factor loading; SMC = squared multiple correlation coefficients. All standardized factor loadings and correlations are significant at $p < 0.001$ (adjusted for the 85 clusters (schools)).

Table A.6: Cronbach's alphas and covariance between PSM dimensions

	Cronbach's alpha	2	3	4
1. Self-sacrifice	0.84	0.31	0.12	0.56
2. Compassion	0.80		0.01	0.38
3. Attraction to public policy making	0.57			0.10
4. Commitment to the public interest	0.70			

Table A.7. Effects of 9th grade teacher PSM on student examination marks: Student fixed effects regressions. Robustness checks: Change of teachers between 7th and 8th grade, and/or between 8th and 9th; average characteristics of teachers in grades 7–9.

Change of teachers	At least 1 change grades 7–9		Change in grades 8–9	
Teacher characteristics	9th grade teachers		7th–9th grade teachers (ave.)	
	(1) Exam marks	(2) Written marks	(3) Exam marks	(4) Written marks
PSM	0.0311*	0.0247*	0.0362***	0.0263**
	(0.0120)	(0.0103)	(0.0106)	(0.00958)
Female teacher	–0.0233	–0.00330	–0.0226	–0.00474
	(0.0297)	(0.0358)	(0.0280)	(0.0326)
Female teacher & student	0.0750**	0.0303	0.0565*	0.0299
	(0.0275)	(0.0354)	(0.0224)	(0.0343)
Qualifications in subject	0.00879	0.00638	0.00309	–0.00289
	(0.0228)	(0.0273)	(0.0229)	(0.0258)
Special teacher edu.	0.0124	0.0449	0.0263	0.0449
	(0.0654)	(0.0660)	(0.0570)	(0.0665)
No teacher education	–0.114	–0.123	–0.0757	–0.165
	(0.115)	(0.194)	(0.129)	(0.190)
Experience 0–4 years	–0.0672*	–0.0867*	–0.0484	–0.0460
	(0.0301)	(0.0388)	(0.0303)	(0.0356)
Experience 5–9 years	0.0340	0.0611*	0.0341	0.0474
	(0.0298)	(0.0306)	(0.0317)	(0.0331)
Change of teachers	–0.00694	0.0286	–0.0134	–0.0114
	(0.0324)	(0.0446)	(0.0433)	(0.0545)
PSM * change of teachers	0.0000772	–0.0218	–0.0201	–0.0480*
	(0.0148)	(0.0179)	(0.0177)	(0.0229)
Observations	24,360	13,660	24,360	13,660
Students	5631	5422	5631	5422
Classes	280	269	280	269
Teachers	694	509	694	509
Schools	85	85	85	85
R^2 (within student)	0.035	0.069	0.034	0.067

All estimates include controls for subjects, interaction terms between teacher characteristics and change of teachers, and interaction terms between dummy variables for subjects and students' gender and immigrant status.
Standard errors in parentheses—robust standard errors clustered on schools.
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$