AARHUS UNIVERSITY

Coversheet

This is the accepted manuscript (post-print version) of the article.

Contentwise, the post-print version is identical to the final published version, but there may be differences in typography and layout.

How to cite this publication

Please cite the final published version:

Lynggard, M., Pedersen, M. J., & Andersen, L. B. (2016). Exploring the Context Dependency of the PSM-Performance Relationship. *Review of Public Personnel Administration*. DOI: 10.1177/0734371X16671371

Publication metadata

Title:	Exploring the Context Dependency of the PSM-Performance Relationship
Author(s):	Lynggard, M., Pedersen, M. J., & Andersen, L. B.
Journal:	<i>Review of Public Personnel Administration</i>
DOI/Link:	http://dx.doi.org/10.1177/0734371X16671371
Document version:	Accepted manuscript (post-print)

General Rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognize and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

EXPLORING THE CONTEXT DEPENDENCY OF THE PSM-PERFORMANCE RELATIONSHIP

Since the early 1990s, scholars have increasingly examined the causes and effects of public service motivation (PSM) (Perry & Hondeghem, 2008; Perry & Wise, 1990), which can be seen as individuals' "orientation to deliver services to people with the purpose of doing good for others and society" (Hondeghem & Perry, 2009, p. 6). Much research shows that employee PSM is a determinant of performance in public organizations (Alonso & Lewis, 2001; Bellé, 2013; Brewer & Selden, 2000; Kim, 2005; Naff & Crum, 1999; Vandenabeele, 2009; Warren & Chen, 2013). However, more research that expands our understanding of the PSM-performance relationship is needed. In particular, scholars call for research attention to the context dependency of the PSM-performance relationship (Bellé, 2013; Pedersen, 2015; Wright & Grant, 2010). How do organizational settings influence the association between PSM and performance in public organizations? What workplace context factors moderate the PSM-performance relationship?

This article contributes by examining how two contextual factors affect the PSM-performance relationship: (1) the extent of work autonomy that public organizations provide employees; and (2) the "user capacity" of the citizens who receive the organization's services, defined as their "competence to understand and affect the provision of the public services" (Kristensen, Andersen, & Pedersen, 2012, p. 947). The work of Le Grand (2003, 2010) and self-determination theory (Gagné & Deci, 2005; Deci & Ryan, 2004) guide our focus. As we discuss later, both perspectives imply that work autonomy and service user capacity may influence the association between PSM and performance. Yet, the two perspectives provide different expectations about the *direction* in which work autonomy and the service user capacity affect the association.

PSM has been shown to be highly relevant in public schools (Van Loon, 2015; Andersen, Heinesen, & Pedersen, 2014). We therefore investigate the context dependency of the association between teacher PSM and student grades. The teachers' work efforts matter for the development of

the individual students and thus for future societal growth and welfare. Because student education serves both the interest of the individual students and the public interest, we focus on the PSM dimensions of “Compassion” and “Commitment to the Public Interest”. Combining survey data from teachers with administrative data on students and schools and using student fixed effects regression, we test how work autonomy and user capacity moderate the PSM-performance relationship in the area of schooling. Do these two contextual factors strengthen or weaken the effect of PSM on performance?

THEORY: THE CONDITIONAL RELATIONSHIP BETWEEN PUBLIC SERVICE MOTIVATION AND PERFORMANCE

PSM theory argues that public sector jobs are intrinsically motivating for people with high PSM, in turn making them work harder and perform better (Perry & Wise, 1990). Empirically, PSM tends to be positively associated with performance in public organizations (Alonso & Lewis, 2001; Bellé, 2013; Brewer & Selden, 2000; Kim, 2005; Naff & Crum, 1999; Vandenabeele, 2009). A meta-analysis by Warren and Chen (2013) confirms the existence of a significant and positive PSM-performance relationship, but they find that association is somewhat weak and stronger in other countries than in the United States. Based on a systematic literature review, Ritz, Brewer and Neumann (2016) suggest that the findings on the PSM-performance relationship are mixed and call for PSM research that do more to integrate context. Examining how contextual factors moderate the PSM-performance relationship is thus a salient issue.

Work Autonomy as a Potential Moderator

The work autonomy given by a public organization to its employees may affect the relationship between the employees’ PSM and performance because employees have more discretion to act on their PSM in organizations with higher work autonomy. The important element of work autonomy to the PSM-performance relationship is the employees’ opportunities to allocate their time to the

work tasks they perceive as the most important. Le Grand (2010) differentiates between two models of public service delivery: trust and mistrust. The *trust* model delegates authority and responsibility to the employees who provide the public services (e.g., nurses, social workers, and teachers) and thus entails high employee work autonomy. The *mistrust* model has “direction from the top, coupled with external rewards or penalties for those complying or failing to comply with the central directives” (Le Grand, 2010, p. 60). The employees are, in other words, not “trusted to do their job properly without outside intervention” (Le Grand, 2010, p. 61). Compared to the trust model, the mistrust model thus gives employees less work autonomy.

Real-life public service is not characterized by ideal-typical trust or mistrust models. Rather, the distinction between trust and mistrust models can be seen as a continuum ranging from low to high employee work autonomy. If PSM is directed towards obtaining the official performance goals, the PSM-performance relationship is expected to be stronger in settings with greater resemblance to a trust model than to a mistrust model, as expressed in hypothesis H1A: *The PSM-performance relationship is stronger in public organizations providing the employees with greater work autonomy.*

The argument is that the trust model results in higher employee performance than the mistrust model because high-PSM employees face greater flexibility and better opportunities to act on their individual public service motives for doing good for others and society (Le Grand 2010). From this perspective, organizational regulation (less work autonomy) is expected to impede the efforts of high-PSM employees when they try to provide the best possible public service.

This logic is consistent with self-determination theory (SDT) (Gagné & Deci, 2005, Deci & Ryan, 2004). SDT distinguishes between amotivation and five types of motivation that differ in the degree to which they are self-determined by the individual (i.e., autonomous and volitional as opposed to controlled or forced by a sense of pressure). From least to most self-determined, the five types of motivation are external, introjected, identified, integrated, and intrinsic motivation (Gagné & Deci, 2005, pp. 333-5). Individual behavior may be based on more controlled types of motivation

(e.g., external or introjected motivation), but support for the basic needs for competence, relatedness, and autonomy may lead to an internalization of the motives for action, in turn resulting in a shift wherein action is guided by a greater extent of motivational self-determination. From a SDT perspective, autonomy involves acting with a sense of volition and having the experience of choice (Gagné and Deci, 2005, pp. 333-5), and more autonomous (self-determined) motivation is expected to promote performance-enhancing behaviors even if these behaviors require self-discipline and effort and are not themselves interesting (Gagné & Deci 2005, p. 339). While PSM does not refer to a fully controlled type of motivation (Pedersen, 2015), having more formal work autonomy supports an individual's basic need for autonomy and may thus promote the self-determination of PSM-oriented behavior (internalizing PSM further)—in turn strengthening the effect of PSM on performance.

However, the expectation that greater work autonomy may increase the strength of the PSM-performance relationship hinges on a key assumption: The high-PSM employees must perceive the performance measure under examination as highly desirable; an ideal way to do good for others and society. In real-life organizations, individuals may have different understandings of what doing good means. If employees prioritize other outcomes than the official performance goals in their efforts to do good, work autonomy might weaken the association between PSM and performance.

While this notion contrasts with the abovementioned expectation (H1A), Le Grand's (2010) theory emphasizes the importance of the employees' own understanding of "doing good". He argues that trust models may not result in high performance if the high-PSM employees are paternalistic, that is, follow their own personal understandings of desirable means and goals. In organizations that provide greater work autonomy, all employees (including those with high PSM) have greater opportunities to pursue individual work motives and goals, and high-PSM employees are expected to have a stronger orientation towards doing good for others and society in terms of their own understanding of what "doing good" means (Gailmard, 2010). Paternalistic employees may, in other words, understand "doing good for others and society" differently from the official

performance goals—and if work autonomy is low, they will have less discretion to follow these alternative understandings. As a result, their PSM may tend to be more directed towards the official understanding of “good” performance when work autonomy is low (creating a stronger PSM-performance association for lower work autonomy). Therefore, a contrasting expectation to H1A is as follows (hypothesis H1B): *The PSM-performance relationship is stronger in public organizations providing the employees with less work autonomy.*

This article uses the variation in employee work autonomy originating in the existence of two different working hour agreements. The main difference between the agreements concerns the teachers’ individual autonomy to allocate their work time to different job tasks (Lynggaard, 2013; Futtrup, 2013). One agreement distributes a fixed number of hours to a set of specific and predefined job task. The other agreement divides the teachers’ working time into only two broad categories, and thus provides teachers with more individual autonomy to allocate their work time to the individual tasks that they themselves believe contribute the most to achieving the goals that they find most important (see the Appendix for more detailed description). Regulation of the allocation of employee working hours entails a difference in work autonomy that may strengthen the PSM-performance relationship for particular aspects of performance because the work efforts of high-PSM employees is directed towards the accomplishment of performance goals that they could otherwise give less priority. This notion is especially salient in the area of schooling; for teacher performance in terms of student exam results. Although achievement of the learning goals in each subject is clearly important, some highly public service-motivated teachers may be inclined to prioritize other goals, for example, relating to their students’ well-being and social development. This is not to say that high-PSM teachers do not acknowledge that student learning is an important school goal. However, they may prioritize other goals that may affect their behavior if they have high autonomy—in turn resulting in a stronger PSM-performance relationship in organizational settings with some regulation of teachers’ work autonomy.

Service User Capacity as a Potential Moderator

Service user capacity—the users’ “competence to understand and affect the provision of the public services” (Kristensen et al., 2012, p. 947)—may moderate the relationship between employee PSM and performance. User capacity varies between users, most importantly because of differences in education. Public services may have both direct users (in schools: students) and indirect users (in schools: parents), and we would like to capture their joint capacity. The capacity to understand and affect school services for a given student is expected to be highly correlated with the capacity of his or her parents. We therefore use a specific parent characteristic—the educational level of a student’s parents—as an indicator for joint service user capacity. Although a direct measure of user capacity would have been preferable to this measure, the literature agrees that capacity relates strongly to education (Eijk & Steen, 2014; Kristensen et al. 2012; Sewell, 1992).

As with work autonomy, we operate with contrasting hypotheses for the way in which the PSM-performance relationship may differ depending on the service user capacity. Andersen and Serritzlew (2012) show that PSM makes employees prioritize weak and disadvantaged users, and Andersen et al. (2014, p. 653) emphasize that norm-based motivation concerning societal contributions is an important part of teachers’ PSM. While “doing good for society” is captured by the PSM dimension “Commitment to the Public Interest,” “Compassion” captures the emotion-based motive for public service emanating from an internal “patriotism of benevolence” towards other people (Perry & Wise, 1990). In schools, students and parents with low service user capacity may find the school experience more difficult, and if highly public service-motivated teachers exert greater work efforts towards helping low user capacity users, we expect that the PSM-performance relationship is stronger when user capacity is weaker. One expectation is thus as follows (hypothesis H2A): *The positive association between teacher PSM and performance is stronger among students with weaker service user capacity.*

The contrasting expectation is, however, also possible (hypothesis H2B): *The positive association between teacher PSM and performance is stronger among students with higher service user capacity.* If high-capacity users seek to improve performance by airing their suggestions about how to accomplish high performance and employees with high PSM want to listen, the combination of PSM and high user capacity may result in higher performance. The effectiveness of voice mechanisms may thus depend both on high service user capacity and high employee PSM (Andersen, Kristensen, & Pedersen, 2013), meaning that teacher PSM may have a stronger effect when the users have a high capacity. Danish schools offer all students and their parents the same formal opportunities to influence service provision, but service user capacity might moderate the PSM-performance relationship positively because high capacity users are better able to use the relevant voice channels. Dialog with service users having the capacity to clearly express their opinions can improve the ability of public service-motivated employees to “do good.”

The Combination of Work Autonomy and User Capacity

In addition to the interactions between PSM and work autonomy and between PSM and user capacity, we examine the existence of three-way interaction between the three variables. The combination of the theoretical mechanisms that we discuss above could, for example, mean that the PSM-performance relationship is stronger in public organizations that combine less work autonomy (making high-PSM employees “do good” as specified by formal performance goals) and low capacity users (inducing public service-motivated employees to act on their PSM even though they have less discretion to follow their own understanding of “doing good”). We do not, however, propose a specific hypothesis regarding the combined interaction effect of work autonomy and user capacity on the PSM-performance relationship because of the contrasting expectations about the theoretical mechanisms.

DATA AND METHODS

Recent research demonstrates the relevance of analyzing the PSM-performance association in a school setting (Van Loon, 2015; Andersen et al., 2014). We may test our expectations in the Danish school sector because teachers differ in their levels of PSM, user capacity varies between schools and between students within schools, and teachers are exposed to one of two working hour agreements entailing different regulation of work autonomy. Moreover, we can measure teacher performance relatively objectively by the students' test score achievements. In addition, this setting allows for user *voice* and *choice*. All parents and students can talk to teachers at formalized meetings or by informal face-to-face dialog. All users can also choose another school than the district school. We can thus focus on moderation of the PSM-performance relationship from varying extents of regulation of employee work autonomy, service user capacity, and the combination of regulation and user capacity. Finally, we are able to use another data set by Andersen et al. (2014) to test the robustness of our findings, allowing us to test our expectations in the same context for two different samples.

Data

We combine survey data and administrative data. The survey data sample consists of Danish school teachers who taught a ninth-grade school class in the subjects Danish or math in 2011. The survey was conducted by The Danish National Centre for Social Research (SFI) and collected in spring 2011. Previous sample analysis gives no reason to reject acceptable sample representativity in terms of school performance, students' average socio-economic status, and school size (Mikkelsen & Lynggaard, 2013). The survey data contains information on teacher variables, including indicators on teachers' PSM and social background characteristics. The administrative data is provided by the National Agency for IT and Learning, an agency under the Danish Ministry for Children, Education and Gender Equality. The data holds information on all Danish lower secondary schools, including the individual students' test scores at the ninth-grade exams.

We link each teacher in the survey with administrative data on each ninth-grade student whom that teacher taught in either Danish or math in the school year of 2010–2011 (both Danish and math teachers can be matched to a specific class and thus to specific students). The resulting data set is structured at the student level with each student represented twice—with one row for Danish matching the Danish teacher to the individual student and another row for math matching the math teacher to that student.

This data structure allows us to test our hypotheses in a panel data setting using student fixed effects (SFE). Compared with typical longitudinal panel data involving two or more observations of the same units over time, we operate with panel data involving two observations of the same students over subjects. While not reaching the same potential for causal inference as randomized experiments, the SFE estimator accounts for subject-invariant confounding at the student, class, manager, and school levels. Our results are therefore relatively robust (see Wooldridge, 2013, pp. 466-83 for a more detailed explanation). Other studies of teacher effects use similar panel setups and SFE estimators (Andersen et al., 2014; Aslam & Kingdon, 2011; Dee, 2007; Pedersen, 2016).

The raw survey data set contains information on 1,642 teachers, but not all teachers completed the questionnaire. Because key indicators were asked in the last part of the survey, incomplete observations were dropped. Moreover, the SFE design entails a matching of each student subject row to the characteristics of an identified subject teacher. All teacher observations that could not be matched to another (opposite-subject) observation within each class were therefore dropped. This approach left data on 316 teachers. The final data set contains 5,676 observations: 2,838 students in 158 classes with 316 teachers in 142 schools.

The reduction in sample size could reduce the representativeness of the sample and introduce selection bias, but sample attrition analyses do not suggest that the 316 teachers are systematically different from the 1,326 teachers who could not be used. We did two-group mean-comparison t-tests in relation to the teacher characteristics of gender, age, education, tenure, teacher experience,

and student performance—and found no differences-in-means (at $p < .05$).¹ Additionally, we provide robustness tests using the data from Andersen et al. (2014) that entail a smaller reduction in sample size because of lacking matches (their survey data comprise all ninth-grade teachers in a school, not Danish and math teachers only, in turn resulting in fewer instances of only a single teacher observation for a given student).

Dependent Variable

We measure performance by the students' ninth grade written exam marks in Danish and math.

High test score achievement is a salient performance goal for all public schools in Denmark.

Because an external examiner takes part in awarding all written exam marks, this performance indicator is relatively objective.² We standardize the students' average scores in Danish and math.

Both measures thus have a mean = 0 and standard deviation = 1.

Independent Variables

Public Service Motivation. We measure the teachers' PSM by a scale comprising seven items covering the PSM dimensions of "Compassion" and "Commitment to the Public Interest" (CPI) (Perry, 1996). These dimensions are especially relevant for performance in schools because they measure the combined orientations towards helping the individual students and towards benefitting society by contributing to the development of future generations of educated citizens.

"Compassion" captures the first-mentioned affective orientation while CPI conceptualizes the last-mentioned normative orientation. We combine them, because "Compassion" and CPI can be seen as

¹ The sample teacher appears to exhibit slightly higher PSM than the teachers that were dropped. This finding is unsurprising: Higher PSM may make people more inclined to help out and respond to a survey request, in turn resulting in higher average PSM among the teachers in the final sample (i.e., comprising both the Danish and the math teacher for a given class). The findings of this article should be viewed in light of this generalizability caveat.

² The external examiner does not know the identity of the individual students. The external examiner and the student's teacher both award a test score. In case of disagreements, the final test score is the mean score with priority to the external examiner score if the mean is not an official grade (see also Andersen, Heinesen, and Pedersen, 2016 for further discussion of the grading institution in Danish schools).

intertwined dimensions for teachers: In a classroom setting, differentiating between “doing good” for the individual students and contributing to society is difficult because what is good for the individual student also contributes to a broader societal interest. The seven items are as follows: (1) “I am often reminded by daily events about how dependent we are on one another”, (2) “I feel emotionally affected when I see people in need”, (3) “To me, considering the welfare of others is one of the most important values”, (4) “I feel that I contribute to society”, (5) “Meaningful public service is very important to me”, (6) “I would prefer seeing public officials do what is best for the whole community even if it harms my interests”, and (7) “I consider public service my civic duty”. The items were translated into Danish with the help of a native English-speaking language editor. Responses were measured using a five-point Likert scale, ranging from 1 (Totally disagree) to 5 (Totally agree). We compute the scale using the mean of the item scores but with the two PSM dimensions given equal weight (adding the mean for the compassion items to the mean for the CPI items and dividing the result by two). Cronbach’s alpha is .79. Confirmatory factor analysis reveals significant factor loadings (at $p < .05$) and close to acceptable goodness-of-fit statistics (CFI = .89; RMSEA = .12; SRMR = .06).

Work Autonomy. We measure work autonomy by exploiting the co-existence of two working hour agreements: A05 and A08. These agreements are important for the teachers’ work autonomy because the agreements allocate (or delegate the allocation of) time to different tasks and thus determine how much discretion each teacher has to decide on the amount of time for different tasks. While the teachers’ total number of working hours is the same under the two agreements, A08 divides the teachers’ working time into two broad categories in which the time allocation is not fixed. A05 regulates the teachers’ working hours more rigidly by dividing the teachers’ working hours into several categories in which the allocation of time is partly fixed. A08 thus gives more autonomy to the teachers in deciding how to spend their working time. The Appendix contains more information about the agreements.

The working hour agreement was made at the municipal level in 2008 by each municipality in negotiation with the local part of the Danish Union of Teachers. All schools within a municipality are subjected to A08 if the two parties managed to reach an agreement. Otherwise, the schools retained the A05 agreement. All schools within a municipality thus operate under the same working hour agreement, and each individual teacher has very limited, if any, influence on the working hour agreement under which he or she is working. An analysis of the municipalities' choice of agreement highlights the role of the decision-makers' personal beliefs about the relative utility of the two agreements and does not find differences related to ideology or use of resources (Futtrup, 2013). From both substantial and methodological perspectives, the teachers' working hour agreement thus serves as a good indicator of work autonomy.³

User Capacity. We measure the joint user capacity (parents and student) for each individual student by the parents' level of education, operationalized as the official standard total time for completion of a given type of education. As mentioned, research suggests that parents' education is an acceptable, although not ideal, indicator of user capacity (Kristensen et al., 2012). The discussion section will discuss the measure further, along with suggestions for future research and measurement.

Controls. We include a rich set of teacher control variables in all model estimations to account for confounding teacher effects (unobserved teacher characteristics are not accounted for by the SFE estimator). Our selection is based on other studies of teacher effects that employ a SFE approach (Andersen et al., 2014; Aslam & Kingdon, 2011; Dee, 2007; Pedersen, 2016). The control variables are as follows: subject (Danish or math teacher), gender, ethnicity (Danish or non-Danish origin), age, education type, pedagogical diploma education, main subject education (in either

³ Our estimates may be biased if teachers with particular characteristics are systematically assigned to one of the working hour agreements. To obtain some leverage on this notion, we conducted a set of t-tests on whether the sample teachers are systematically distinct across the A05 and A08 working hour agreements in terms of teacher gender, ethnicity, age, education, tenure, fulltime employment, and PSM. We find no significant differences-in-means ($p < .05$).

Danish or math), number of years at school, number of years as teacher in the class, and fulltime employment.

The main variables of the article are measured at different levels: Performance and service user capacity are measured at the student level; PSM is measured at the teacher level; and extent of work autonomy is measured at the school level.⁴ We now elaborate on the SFE approach that we use to account for this multilevel data structure.

Estimation Strategy

The PSM-performance relationship can be estimated by a basic production function:

$$(eq. 1) \quad Y_{ijk} = \beta_{0j} + U'_{ijk}\beta_{1j} + T'_{ijk}\beta_{2j} + S'_{ik}\beta_{3j} + W'_{ijk}\beta_{4j} + \epsilon_{ijk}$$

Y_{ijk} represents the performance of student i in subject j at school k . Student performance is a function of student characteristics (U), teacher characteristics (T), and school characteristics (S), while W represents the effect of teacher PSM. Estimating eq. (1) by basic OLS regression produces biased estimates if any unobserved student, teacher, or school characteristics occur systematically in the error term (ϵ) and are correlated with the explanatory variables. To account for such potential bias sources, this article employs a SFE estimator. By comparing the same student's Danish and math performances, the effect of all subject-invariant student and school characteristics is eliminated. Such between-subject differencing implies that:

$$(eq. 2) \quad \Delta Y_i = (\beta_{0,d} - \beta_{0,m}) + U'_i(\beta_{1,d} - \beta_{1,m}) + T'_{i,d}\beta_{2,d} - T'_{i,m}\beta_{2,m} \\ + S_i(\beta_{3,d} - \beta_{3,m}) + W'_{i,d}\beta_{4,d} - W'_{i,m}\beta_{4,m} + \eta_i$$

⁴ We have conducted a set of tests that examine the robustness of our findings when aggregating our variable measurements to higher-order levels. We have retested all statistical models using service user capacity measures at the class and school levels (i.e., average service user capacity for each school class and each school). We have employed binary measures of user capacity (at the student, school class, and school levels). We have also estimated class fixed effects models that aggregate all student level measures (including performance) to the teacher level. Importantly, none of these tests result in any substantive changes to our main findings.

Subscript d (m) relates an observation to the subject Danish (math). ΔY_i denotes the within-difference in performances of student i (i.e., $Y_{id} - Y_{im}$) while η_i denotes the error term. Assuming that student and school characteristics have the same effect on student performance across subjects (that school and student influences on performance are subject unspecific) means that $\beta_{1,m} - \beta_{1,d} = \beta_{3,m} - \beta_{3,d} = 0$. These effects are therefore removable from eq. (2). With $\delta = \beta_{0,m} - \beta_{0,d}$, eq. (2) is thus equivalent to:

$$(eq. 3) \quad \Delta Y_i = \delta + \Delta T'_i \beta_2 + \Delta W'_i \beta_4 + \eta_i$$

Eq. (3) represents the SFE estimation approach. Only teacher characteristics (T), including their PSM (W), are retained because these are the only characteristics varying within students. The SFE estimation of PSM coefficients by eq. (3) is unbiased by subject-invariant student, class, manager, and school effects. We present both OLS and SFE fixed effect regressions to allow for comparison between estimation strategies.

RESULTS

Table 1 presents descriptive statistics of the measures that we include in the statistical models. As all student and school characteristics are controlled for in the SFE regressions, student and school variables are only included in the basic OLS regressions. The differences in the number of observations reflect the level of analysis at which we measure the individual variables. For example, performance and user capacity are measured at the student level, PSM at the teacher level, and working hour agreement at the school level.

Insert Table 1 about here

Table 2 displays the main results. Models 1A and 1B show the PSM-performance relationship without any interaction terms. Models 2A and 2B include an interaction term for working hour agreement A08 and PSM, thus testing whether work autonomy moderates the PSM-performance

relationship. Models 3A and 3B include an interaction term for user capacity and PSM, thus testing whether service user capacity moderates the PSM-performance relationship. Finally, models 4A and 4B include interaction terms capturing the three-way interaction between PSM, work autonomy, and service user capacity.⁵ The “A” models show the results of OLS regressions, including student and school controls; the “B” models show the results of SFE regressions, which are similar except that student and school controls are even better given that we only use the variation between the same student’s different subjects where student and school variables do not vary. Standard errors are clustered at the school level to account for inter-cluster correlation produced by a nesting of teachers in schools. This is a conservative test, also given that measurement errors tend to be exacerbated in fixed effect models, potentially biasing the results towards zero. Therefore (and also considering that Warren and Chen (2013) describe the PSM-performance association as weak), the insignificant coefficient for the main effect of PSM should not be used to rule out that PSM affects performance.

Insert Table 2 about here

Higher user capacity is associated with higher performance in all models. This result is unsurprising as we operationalize user capacity by the educational level of the students’ parents. The SFE models (1B, 2B, 3B, and 4B) account for this effect by comparing only between a given student’s grades in different subjects. The same is the case for work autonomy (dummy variable, $A08 = 1$), which is constant within schools and therefore does not vary between subjects for the individual student.

⁵ As a robustness test, we reran model 4B three times (substituting, first, our PSM measure with a sum scale capturing “compassion,” then using a sum scale capturing “commitment to the public interest,” and, finally, using a combined PSM measure that we constructed using the predicted factor scores). The results are qualitatively identical to those presented in Table 2, model 4B.

Models 1A and 1B show a positive PSM coefficient, which indicates that higher teacher PSM relates to higher performance. The estimates are not statistically significant. The interaction coefficients in models 2A and 2B are negative, which suggests that the PSM-performance relationship is stronger in work settings with some regulation of teacher work autonomy. The PSM-performance association is 0.19 for schools under the working hour agreement with more regulation (A05) while it approximates zero for the A08 working hour agreement (the negative interaction term coefficients cancel out the positive PSM coefficient).

Models 3A and 3B test for interaction effects between PSM and service user capacity. The findings indicate that service user capacity does not moderate the PSM-performance relationship (the coefficient for the PSM-user capacity interaction term is close to zero and insignificant). This null-finding does not, however, mean that service user capacity is not relevant for the PSM-performance relationship. Model 4B thus shows a significant three-way interaction effect between PSM, working hour agreement, and service user capacity. Under the working agreement with relatively more regulation (A05), the combination of high PSM and low user capacity has a positive association with performance. To illustrate this finding, we use the model 4B estimates to compute the marginal effects of PSM on performance for different models of public service provision (A05/A08) and service user capacity. The dotted line in Figure 1 shows the marginal effects of a one-unit increase in PSM on performance at different levels of service user capacity for schools with more restricted teacher work autonomy (A05). The fully drawn line shows the marginal effects for schools under A08. The vertical lines represent 95 percent confidence bands at the different levels of user capacity.

Insert Figure 1 about here

In support of H1B and H2A, Figure 1 shows that the PSM-performance relationship is significant, but only for schools with greater extent of regulation (A05) and for students having

parents with low to moderate user capacity. Please note that performance is conceptualized as formal goal attainment and measured as exam grades in Danish and math. Other understandings of what “doing good for others and society” means are possible, and the public service-motivated teachers might prioritize these and have the autonomy to work towards them if work autonomy is high.

In sum, our results suggest a *conditional* positive association between PSM and performance, such that higher PSM is related to higher performance but with varying strength. The finding implies that some level of regulation may yield a formal structuring of the direction and focus of the employees’ work behaviors—a structuring appearing beneficial to the employees’ performance invoked by PSM. We discuss how to interpret this finding after the presentation of our key robustness tests.

Andersen et al. (2014) have examined the PSM-performance relationship for other Danish schools. We use their data to test the replicability of our results. Andersen et al.’s (2014) sample includes fewer schools (85 schools) but more teachers per school and more performance indicators per student (grades in all subjects rather than only Danish and math). Also, their measure of PSM includes all four dimensions specified by Perry (1996). Despite these differences, the results are substantially similar across data sets (see the Appendix, Table A1). This consistency of findings persists regardless of whether all exams or only written exams are used and whether we look at students graduating in 2011 (when the exams took place after the measurement of PSM) or all students graduating between 2009 and 2011. This robustness check suggests that our results are not due to stochastic variations in the sample investigated in this article. Before the conclusion sums up the findings and their implications, the next section presents possible explanations of our key finding and discusses the study’s limitations.

DISCUSSION

This section, first, discusses why PSM affects performance more positively under a working hour agreement involving lesser (as opposed to greater) employee work autonomy – especially for service users with a low to moderate service user capacity. We then discuss three methodological issues, namely the potential for causal inference given the SFE estimation and our measures of user capacity and PSM.

PSM seems to have the strongest effect on performance when work autonomy is limited and user capacity is low to moderate. This finding may be surprising, but existing research contributes to an understanding of this key finding. While greater extent of employee work autonomy may offer better opportunities for realization of public service motives to do good for society and others, it also allows high-PSM individuals to pursue their own understandings of “doing good for society and others”. In the absence of strong regulation, PSM may increase only the aspects of performance that high-PSM employees prioritize, and these aspects might differ from the official performance goals. Recent research thus emphasizes that different stakeholders can have very different performance criteria (Andersen, Boesen, & Pedersen, 2016). If teachers believe that students’ test scores are important but that other goals are more desirable, public service-motivated employees may follow these other goals like “runaway agents” (Gailmard, 2010; Kiewiet & McCubbins, 1991). This notion may explain why the positive relationship between PSM and (officially defined) performance is more pronounced in settings with some regulation (less work autonomy). The identified three-way interaction with user capacity might be explained by the fact that public service-motivated employees are especially oriented towards helping the weaker users, thus making them put in more effort despite the stronger regulation in a context with relatively low work autonomy.

Although we see the above interpretation as the most plausible, there are other potential theoretical explanations. For example, greater autonomy given to each teacher to administer his or her working hours (Kamp et al., 2011) can decrease the maneuverability for both municipalities and school principals because the teachers (and not the municipalities or principals) have the right to

allocate time to specific tasks. This may diminish managerial responsiveness, cohesion, and buffering capacities—which, in turn, may suppress and counteract positive effects of teacher PSM on performance under the working hour agreement with relatively more autonomy. This maneuverability might be especially important in contexts with low-capacity users, explaining the three-way interaction.

Our findings contrast with the SDT-based expectation that greater work autonomy increases the strength of the PSM-performance relationship. Rather, our findings are in line with Le Grand's (2010) theory and notion that the employees' personal understandings of what “doing good” means may condition the moderating role of work autonomy. In settings where high-PSM employees' own understandings may differ from the official performance goals, lower work autonomy may result in a more positive relationship between PSM and the official performance goals. In SDT terminology, support for the basic need for autonomy may thus result in a greater internalization of individuals' motives for action, but the resulting self-determined behavior aimed at doing good for others and society may only contribute to reaching performance goals that the employees themselves prioritize.

Moreover, our study is not exempted from methodological caveats and limitations. The SFE estimates are relatively robust, and we replicate the findings using another sample, but four aspects are worth discussing in more detail: remaining challenges to causal inference and our measures of user capacity, work autonomy and PSM. Most importantly, we recognize that the SFE does not automatically allow for causal conclusions. Confounding caused by unobserved teacher effects is still a risk, and reverse causality bias is also a potential threat. Still, both the main analysis and the robustness test control for the most relevant teacher characteristics, and we find no evidence suggesting that student characteristics vary systematically between subjects. Both surveys were collected before the 2011 exam, and sorting bias (e.g., letting the best teacher teach the most difficult class) is controlled for in the SFE as long as the selection is not subject specific (because students attend the same classes in all subjects). Finally, our use of separate data sources for all key variables prevents common source bias. In sum, our research design and methods contribute to

make the findings relatively robust, but we hope that future research will test our findings using research designs that provide even better opportunities for causal inference.

A main limitation of the article is our measure of user capacity. For user capacity, our indicator is the parents' education. This crude operationalization reflects that public administration research only recently began to focus on this important variable, responding to Le Grand's (2010) arguments. The data for our study was already collected when the first direct measure of perceived user capacity was published (Kristensen et al., 2012). However, parents' education may be an acceptable (although not ideal) indicator of user capacity in a school context for several reasons. First, the empirical association between education and user capacity is significant and strong (Kristensen et al., 2012). Second, theory supports that cultural capital (and thus also education) is likely crucial for users' potential for affecting service provision in educational institutions such as schools (e.g., Bourdieu & Passeron, 1990). Third, parents normally complete their education before their child starts school, thus limiting potential endogeneity in this part of the model.

With parents' education as a measure of user capacity, our results suggest that user capacity is an important contextual factor for the PSM-performance relationship. Discussing future conceptualization and measurement of user capacity is therefore relevant. While we argue that the use of parents' education can be acceptable, Kristensen et al. (2012) have developed a direct measure of perceived user capacity, and future studies could benefit from using both an objective measure (such as education) and a subjective measure. Users' actions may depend both on their objective capacities and on their own perceived ability to influence public service provision. In addition, measuring the capacity of the direct and the indirect users separately could be useful. In this article, we argue that parents' education is an indicator for the joint user capacity of parents and students. While this is plausible in a school context, the use of separate measures may be important in other contexts. Investigations of elder care could, for example, differentiate between the capacity of direct users (the elderly) and of indirect users (next of kin). In sum, we strongly recommend that future research continue the efforts to include better measures of user capacity.

Moreover, our measure of work autonomy relates to the freedom to allocate time at work to predefined job tasks. As mentioned by an anonymous reviewer, however, work autonomy may also involve the freedom to choose the particular work tasks you want to perform yourself. Our findings should be interpreted and extrapolated in perspective of this limitation.

Finally, our measure of PSM is not beyond discussion. Importantly, however, our robustness tests use a full version of Perry's (1996) conceptualization and show the same results. Moreover, we agree with Ritz et al. (2016, p. 423), who state that which of the existing PSM scales has the best measurement properties is unclear. We hope that future research continues to improve the PSM measure, but fortunately, different measures provide almost identical results in our study.

CONCLUSION

This article offers new knowledge on how contextual factors may moderate the PSM-performance relationship. We find that the PSM-performance relationship is moderated by the extent of work autonomy that a public organization provides to its employees. In particular, PSM appears to affect performance more positively under a working hour agreement involving lesser (as opposed to greater) employee work autonomy—especially for service users with a low to moderate service user capacity.

Our findings contribute to the PSM literature by expanding our understanding of the PSM-performance relationship. By focusing on two organizational context factors (work autonomy and service user capacity), we illustrate that the PSM-performance relationship can be context dependent; more pronounced in some workplace settings than in others. We recognize that a range of other context variables may affect the PSM-performance relationship, and future research should continue the examination of these context effects. For example, Kjeldsen (2012) distinguishes between public organizations that deliver “core public services” and those that do not, and Van Loon (2015) finds that PSM is systematically related to performance-related behaviors in people-changing organizations (service production aimed at changing the user, such as schools) while there

is no association between PSM and output and responsiveness in people-processing organizations (service regulation categorizing and processing users, such as police departments).

Future research will hopefully also improve internal and external validity of this article's findings. For example, the results could be tested in other contexts and by experimental or quasi-experimental methods. In terms of generalization to other empirical contexts than Danish schools, we have no reason to expect the investigated context to differ from many other public service organizations, but appliance of the conceptual framework in another institutional setting would be very helpful. Especially for future studies investigating user capacity in other types of organizations than schools, we recommend the use of better measures of user capacity.

Greater attention to the context dependency of the PSM-performance relationship is not only of scholarly interest. Knowledge on when and where the PSM-performance relationship is especially pronounced is also important to practice, for example, if public managers are to increase performance by increasing or engaging their employees' PSM. As the implementation of new organizational practices is neither easy nor inexpensive, public managers will benefit from knowledge on the specific workplace settings wherein employee PSM is especially stimulating for high performance.

REFERENCES

- Alonso, P., & Lewis, G. B. (2001). Public service motivation and job performance: Evidence from the federal sector. *American Review of Public Administration*, 31(4), 363-380.
- Andersen, L. B., Boesen, A., & Pedersen, L. H. (2016). Performance in public organizations: Clarifying the conceptual space. Forthcoming in *Public Administration Review*.
- Andersen, L. B., Heinesen, E., & Pedersen, L. H. (2014). How does public service motivation among teachers affect student performance in schools? *Journal of Public Administration Research and Theory*, 24(3), 651-671.

- Andersen, L. B., Heinesen, E., & Pedersen, L. H. (2016). Individual performance: From common source bias to institutionalized assessment. *Journal of Public Administration Research and Theory*, 26(1), 63-78.
- Andersen, L. B., Kristensen, N., & Pedersen, L. H. (2013). Models of public service provision—When will knights and knaves be responsive to pawns and queens? *International Journal of Public Administration*, 36(2), 126-136.
- Andersen, L. B., & Serritzlew, S. (2012). Does public service motivation affect the behavior of professionals? *International Journal of Public Administration*, 35(1), 19-29.
- Aslam, M., & Kingdon, G. (2011). What can teachers do to raise pupil achievement? *Economics of Education Review*, 30(3), 559-574.
- Bellé, N. (2013). Experimental evidence on the relationship between public service motivation and job performance. *Public Administration Review*, 73(1), 143-153.
- Bourdieu, P., & Passeron, J. C. (1990/1977). *Reproduction in education, society and culture*, London, Thousand Oaks, New Delhi: Sage Publications Inc.
- Brewer, G. A., & Selden, S. C. (2000). Why elephants gallop: Assessing and predicting organizational performance in federal agencies. *Journal of Public Administration Research and Theory*, 10(4), 685-711.
- Deci, E. L., & Ryan, R. M. (2004). *Handbook of self-determination research*, Rochester: University of Rochester Press.
- Dee, T. S. (2007). Teachers and the gender gap in student achievement. *Journal of Human Resource*, 42(3), 528-554.
- Eijk, C. J. A. van, & Steen, T. P. S. (2014). Why people co-produce: Analysing citizens' perceptions on co-planning engagement in health care services, *Public Management Review*, 16(3), 358-382.

- Futtrup, R. B. (2013). *Variationer i kommunal bureaukratisering: Hvorfor kommunerne anvender forskellig detaljeringsgrad i lærernes arbejdstidsaftale*.
<http://www.specialer.sam.au.dk/stat/2013/20071045.pdf>.
- Gagné, M., & Deci, E. L. (2005). Self-determination theory and work motivation. *Journal of Organizational Behavior*, 26(4), 331-362.
- Gailmard, S. (2010). Politics, principal-agent problems, and public service motivation. *International Public Management Journal*, 13(1), 35-45.
- Hondeghem, A., & Perry, J. L. (2009). EGPA symposium on public service motivation and performance: Introduction. *International Review of Administrative Sciences*, 75, 5-9.
- Kamp, A., Lund, H., Holt, H., & Hvid, H. (2011). Retten til egen tid [The right to administer your own time]. *Tidsskrift for Arbejdsliv*, 1, 15-30.
- Kim, S. (2005). Individual-level factors and organizational performance in government organizations. *Journal of Public Administration Research and Theory*, 15(2), 245-262.
- Kiewiet, D. R., & McCubbins, M. D. (1991). *The logic of delegation*. Chicago: University of Chicago Press.
- Kjeldsen, A. M. (2012). Vocational study and public service motivation: Disentangling the socializing effects of higher education. *International Public Management Journal*, 15(4), 500-524.
- Kristensen, N., Andersen, L. B., & Pedersen, L. H. (2012). Public service efficacy. *International Journal of Public Administration*, 35(14), 947-958.
- Le Grand, J. (2003). *Motivation, agency and public policy*. Oxford: Oxford University Press.
- Le Grand, J. (2010). Knights and knaves return: Public service motivation and the delivery of public services. *International Public Management Journal*, 13(1), 56-71.
- Lynggaard, M. (2013). Lærernes arbejdstidsaftaler [The teachers' working hour agreements]. In *Lærere, undervisning og elevpræstationer i folkeskolen* [Teachers, teaching, and student performance in the primary and lower secondary school], edited by Winter, S. C., &

- Nielsen, V. L., 243-260. Copenhagen: SFI – The Danish National Centre for Social Research.
- Mikkelsen, M. F., & Lynggaard, M. (2013). Datamateriale og metoder [Data and methods]. In *Lærere, undervisning og elevpræstationer i folkeskolen* [Teachers, teaching, and student performance in the primary and lower secondary school], edited by Winter, S. C., & Nielsen, V. L., 175-190. Copenhagen: SFI – The Danish National Centre for Social Research.
- Naff, K. C., & Crum, J. (1999). Working for America: Does public service motivation make a difference? *Review of Public Personnel Administration*, 19(4), 5-16.
- Pedersen, M. J. (2016). A “Heart of goal” and the will to succeed: Goal commitment and task performance among teachers in public schools. *Public Administration*, 94(1), 75-88.
- Pedersen, M. J. (2015). Activating the forces of public service motivation: Evidence from a low-intensity randomized survey experiment. *Public Administration Review*, 75(5), 734-746.
- Perry, J. L. (1996). Measuring public service motivation: An assessment of construct reliability and validity. *Journal of Public Administration Research and Theory*, 6(1), 5-23.
- Perry, J. L., & Hondeghem, A. (2008). Building theory and empirical evidence about public service motivation. *International Public Management Journal*, 11(1), 3-12.
- Perry, J. L., & Wise, L. R. (1990). The motivational bases of public service. *Public Administration Review*, 50(3), 367-373.
- Ritz, A., Brewer, G. A., & Neumann, O. (2016). Public service motivation: A systematic literature review and outlook. *Public Administration Review*, 76(3), 414-426.
- Sewell, W. H. J. (1992). A theory of structure: Duality, agency and transformation. *American Journal of Sociology*, 98(1), 1-29.
- Van Loon, N. M. (2015). Does context matter for the type of performance-related behavior of public service motivated employees? *Review of Public Personnel Administration*. Published online before print, June 17, 2015.

- Vandenabeele, W. (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), 11-34.
- Warren, D. C., & Chen, L. (2013). Meta-analysis of the relationship between public service motivation and performance. In *Meta-analysis for public management and policy*, edited by Evan J. Ringquist, 442-474. San Francisco, CA: Jossey-Bass.
- Wooldridge, J. M. (2013). *Introductory econometrics: A modern approach (5e)*. Cincinnati, OH: South-Western College Publishing.
- Wright, B. E., & Grant, A. M. (2010). Unanswered questions about public service motivation: Designing research to address key issues of emergence and effects. *Public Administration Review*, 70(5), 691-700.

Table 1 Descriptive Statistics for Study Variables

	Obs.	Mean	S.D.	Min	Max
<i>Main variables</i>					
Student grades	5676	.00	1.00	-2.46	2.25
Service user capacity	5676	12.68	2.06	0	20
PSM	316	4.09	.53	2	5
A08	142	.85	.36	0	1
<i>Control variables: Student</i>					
Gender	5676	.51	.50	0	1
Ethnicity	5674	.08	.27	0	1
Age at time of test	5676	16.11	.39	14.92	17.92
Birth Weight > 2500g	5368	.95	.21	0	1
Sibling status: Only Child	5676	.07	.26	0	1
----- : Firstborn	5676	.37	.48	0	1
----- : Middle born	5676	.17	.38	0	1
----- : Last born	5676	.39	.49	0	1
Age at student's birth, mother	5516	29.06	4.60	16	45
Age at student's birth, father	5378	31.66	5.50	18	59
Family structure: Does not live with parents	5676	.00	.06	0	1
----- : Lives with a single parent	5676	.12	.32	0	1
----- : Lives with parent and his/her new partner	5676	.08	.26	0	1
----- : Lives in a nuclear family	5676	.80	.40	0	1
Disp. income (log), mother	5670	12.19	1.04	0	13.74
Disp. income (log), father	5652	11.95	2.20	0	14.67
Employment status, mother: Employed	5676	.86	.35	0	1
----- : Unemployed	5676	.07	.26	0	1
----- : Attending education	5676	.00	.06	0	1
----- : Pension/outside labor force	5676	.04	.20	0	1
Employment status, father: Employed	5676	.86	.34	0	1
----- : Unemployed	5676	.06	.24	0	1
----- : Attending education	5676	.00	.02	0	1
----- : Pension/outside labor force	5676	.04	.20	0	1
<i>Control variables: Teacher</i>					
Subject	316	.5	.5	0	1
Gender	308	.56	.50	0	1
Ethnicity	309	.95	.21	0	1
Age	308	46.77	10.67	25	66
Education type	303	.07	.26	0	1
Diploma education	266	.14	.34	0	1
Main subject	308	.75	.44	0	1
Years at school	307	12.86	10.98	0	43
Years as class teacher	306	2.95	1.67	0	9
Fulltime employed	315	.93	.25	0	1
<i>Control variables: School</i>					
School size	142	473.09	162.53	123	931
Proportion of 9 th grade non-Danish students	142	.11	.14	0	1
9 th grade students' parents' length of education	142	12.59	.80	9.90	15.09
Proportion of 9 th grade students' mothers in highest income quartile	142	.33	.15	0	.75
Proportion of 9 th grade students' fathers in highest income quartile	142	.28	.15	0	0.81

Table 2 Regressions Analyzing the PSM-Performance Relationship. Student Grades as the Dependent Variable

	PSM		PSM x A08		PSM x user capacity		PSM x A08 x user capacity	
	OLS (1A)	SFE (1B)	OLS (2A)	SFE (2B)	OLS (3A)	SFE (3B)	OLS (4A)	SFE (4B)
PSM for teacher in given subject	.06 (.03)	.01 (.03)	.19*** (.05)	.19** (.06)	.06 (.16)	.01 (.13)	.65* (.26)	1.11*** (.17)
A08 [†]	.06 (.06)	-	.69** (.25)	-	.06 (.06)	-	3.15* (1.27)	-
Service user capacity [†]	.13*** (.01)	-	.13*** (.01)	-	.13* (.05)	-	.30** (.08)	-
PSM x A08	-	-	-.16* (.06)	-.22** (.07)	-	-	-.71* (.31)	-1.28*** (.21)
PSM x service user capacity	-	-	-	-	.00 (.01)	-.00 (.01)	-.04 (.02)	-.07*** (.01)
A08 x service user capacity [†]	-	-	-	-	-	-	-.20 (.10)	-
PSM x A08 x service user capacity	-	-	-	-	-	-	.04 (.03)	.09*** (.02)
Teacher controls:								
Gender	.04 (.04)	.07* (.04)	.04 (.04)	.07 (.04)	.04 (.04)	.07* (.04)	.04 (.04)	.07 (.04)
Ethnicity	.10 (.08)	-.06 (.08)	.11 (.08)	-.07 (.07)	.10 (.08)	-.06 (.08)	.11 (.08)	-.07 (.08)
Age	-.00 (.00)	.00 (.00)	-.00 (.00)	.00 (.00)	-.00 (.00)	.00 (.00)	-.00 (.00)	.00 (.00)
Higher education	.03 (.06)	.04 (.07)	.02 (.06)	.03 (.07)	.03 (.06)	.04 (.07)	.02 (.06)	.03 (.07)
Diploma education	.01 (.05)	-.02 (.06)	.01 (.05)	-.02 (.06)	.01 (.05)	-.02 (.06)	.01 (.05)	-.02 (.06)
Main subject	-.02 (.04)	-.02 (.04)	-.01 (.04)	-.02 (.04)	-.02 (.04)	-.02 (.04)	-.01 (.04)	-.02 (.04)
Years at school	.00 (.00)	-.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)	-.00 (.00)	.00 (.00)	.00 (.00)
Years as class teacher	.00 (.01)	.01 (.01)	.00 (.01)	.01 (.01)	.00 (.01)	.01 (.01)	.00 (.01)	.01 (.01)
Fulltime employed	-.07 (.06)	-.11 (.07)	-.07 (.06)	-.12 (.07)	-.07 (.06)	-.11 (.07)	-.07 (.06)	-.12 (.07)
Student fixed effects	No	Yes	No	Yes	No	Yes	No	Yes
Student and school controls	Yes	No	Yes	No	Yes	No	Yes	No
Observations (no.)	5,676	5,676	5,676	5,676	5,676	5,676	5,676	5,676
Groups (no.)	-	2,838	-	2,838	-	2,838	-	2,838
Adj. R ²	.20	.02	.20	.02	.20	.02	.20	.03

Notes: * $p < .05$; ** $p < .01$; *** $p < .001$. [†]: These variables drop out of the SFE models, as these characteristics are constant for the individual student. Clustered standard errors (school level) in parentheses. All regressions include a subject dummy. “Adj. R²” refers to the “within” estimates for all SFE models. For brevity, we do not report the coefficients for the student and school control variables included in the “OLS” models. The full models are available from the authors.

Figure 1 Marginal Effects Plot, PSM on Performance

APPENDIX

Box A1: Description of the Two Working Hour Agreements

Both working hour agreements (A05 and A08) were made between the teachers' union and employer representatives (the national advocacy organization of Danish municipalities, in English, "Local Government Denmark"). The names of the working hour agreements refer to the year when the agreements were last revised or adopted (2005 and 2008). As mentioned, the main difference between them concerns the teachers' individual autonomy to allocate their work time to different tasks (Lynggaard, 2013; Futtrup, 2013). A05 distributes a fixed number of hours to a set of specific and predefined tasks; for example, 375 hours to the teachers' individual preparation and post treatment of their teaching, 155 hours to the teachers' cooperation, teamwork, and tasks related to school development, and 75 hours to the task of being responsible for a given school class, including contact with students' parents. Additionally, A05 provides recommendations and guidelines about how school principals should allocate the teachers' remaining hours in relation to other tasks. In contrast, A08 divides the teachers' working time into only two broad categories, named "the teaching task" and "other tasks," in which the time allocation is not fixed across the two categories.

Futtrup (2013) analyses the variation in the municipalities' choice of working hour agreements and does not find that political factors matter. The decision-makers state that maintaining cost efficiency was important, and the resources per student did not change when municipalities shifted from one agreement to the other. The decision seems to depend on the decision-makers' personal beliefs about what contributes the most to student learning (Futtrup, 2013, p. 90). As our sample teachers did not take active part in these decisions, the risk of this particular endogeneity problem is very small.

Table A1. Robustness Check Using Data from Andersen, Heinesen, and Pedersen (2014). Student Fixed Effects.

	2009-2011		2011	
	Exam Marks (Model 1)	Written Marks (Model 2)	Exam Marks (Model 3)	Written Marks (Model 4)
PSM for teacher in given subject	.12** (.03)	.10*** (.03)	.15** (.04)	.27** (.10)
PSM x A08	-.10* (.04)	-.09* (.04)	-.14* (.06)	-.26* (.10)
PSM x service user capacity	-.00 (.00)	-.01*** (.00)	-.00 (.00)	-.02* (.01)
PSM x A08 x service user capacity	.00 (.00)	.01* (.00)	.01 (.00)	.02* (.01)
Teacher controls:				
Gender	.01 (.02)	.01 (.02)	-.02 (.03)	.01 (.03)
Female teacher and student	.03* (.02)	.01 (.02)	.04 (.02)	.01 (.04)
Qualifications in subject	.01 (.01)	.00 (.02)	.00 (.02)	-.00 (.03)
Special teacher education	-.01 (.03)	.03 (.03)	.05 (.05)	.07 (.06)
No teacher education	-.16* (.07)	-.14 (.09)	-.07 (.11)	-.14 (.16)
Experience, 0-4 years	-.04* (.02)	-.01 (.02)	-.04 (.03)	-.05 (.04)
Experience, 5-9 years	.02 (.02)	.02 (.02)	.03 (.03)	.00 (.03)
Shift of teachers	.00 (.01)	-.02 (.02)	-.05 (.03)	-.04 (.03)
Observations (no.)	53,002	30,031	20,263	11,345
R ²	0.030	0.054	0.036	0.069

Notes: * $p < .05$; ** $p < .01$; *** $p < .001$. Clustered standard errors (school level) in parentheses. All regressions include a subject dummy. “R²” refers to the “within” estimates. Interactions between subject and immigrant status as well as interactions between student gender and subjects are included, but not shown. The full models are available from the authors.