Gjefsen, Hege Marie; Gunnes, Trude

Working Paper

School accountability: Incentives or sorting?

Discussion Papers, No. 815

Provided in Cooperation with:
Research Department, Statistics Norway, Oslo

Suggested Citation: Gjefsen, Hege Marie; Gunnes, Trude (2015) : School accountability: Incentives or sorting?, Discussion Papers, No. 815, Statistics Norway, Research Department, Oslo

This Version is available at:
https://hdl.handle.net/10419/192797

Discussion Papers
Statistics Norway
Research department

No. 815
July 2015

Statistics Norway

Statistisk sentralbyrå

Hege Marie Gjefsen and Trude Gunnes

School accountability: Incentives or sorting?

Discussion Papers No. 815, July 2015
Statistics Norway, Research Department

Hege Marie Gjefsen and Trude Gunnes

School accountability: Incentives or sorting?

Abstract:

We exploit a nested school accountability reform to estimate the causal effect on teacher mobility, sorting, and student achievement. In 2003, lower-secondary schools in Oslo became accountable to the school district authority for student achievement. In 2005, information on school performance in lower secondary education also became public. Using a difference-in-difference-in-difference approach, we find a significant increase in teacher mobility and that almost all non-stayers leave the teaching sector entirely. The impact is larger on high-ability teachers following the second part of the reform. Non-stayers are largely replaced by high-ability teachers, indicating a positive sorting effect. We find a small, positive effect on student achievement after the second part of the reform, thus the mechanism in place seems to be positive teacher sorting rather than teacher incentives.**Keywords:** school accountability regimes, design of incentives, teacher turnover, teacher quality, teacher sorting, student achievement, difference-in-difference-in-difference

Keywords school accountability regimes, design of incentives, teacher turnover, teacher quality, teacher sorting, student achievement, difference-in-difference-in-difference

JEL classification: D03, I21, J21, J38

Acknowledgements: We are grateful for generous comments from Lars-Erik Borge, Torberg Falch, David Figlio, Lars Kirkebøen, Bjarne Strøm, Kjetil Telle, in addition to comments from seminar participants at the Norwegian annual meeting for economists in Oslo 2014, the Royal Economic Society Conference in Manchester 2014, and the SOLE/EALE world meeting in Montreal 2015. We are also grateful for data provided by the Norwegian Universities and Colleges Admission Service and for valuable comments from the school district authority in Oslo.

Address: Trude Gunnes, Statistics Norway, Research Department. E-mail: gut@ssb.no

Hege Marie Gjefsen, Statistics Norway, Research Department. E-mail: hmg@ssb.no

Discussion Papers	comprise research papers intended for international journals or books. A preprint of a Discussion Paper may be longer and more elaborate than a standard journal article, as it may include intermediate calculations and background material etc.

Abstracts with downloadable Discussion Papers
in PDF are available on the Internet:
http://www.ssb.no/en/forskning/discussion-papers
http://ideas.repec.org/s/ssb/dispap.html

ISSN 1892-753X (electronic)

Sammendrag

Vi utnytter en todelt mål- og resultatstyringsreform i Oslo-skolen for å estimere den kausale effekten på lærermobilitet og elevprestasjoner. I 2003 ble ungdomsskoler i Oslo ansvarliggjort for sine elevprestasjoner opp mot Utdanningsetaten. I 2005 ble også skoleprestasjoner offentliggjort. Ved å bruke en forskjeller-i-forskjeller-i-forskjeller metode, finner vi en signifikant positiv effekt på lærermobilitet etter begge reformene. Lærere som bytter jobb forlater undervisningssektoren. Mobiliteten er størst blant kvalitetslærere (målt ved lærernes egne akademiske resultater) etter den andre delen av reformen. Tilsiget av nye lærere i Oslo er også av høy kvalitet, noe som indikerer en positiv seleksjonseffekt. Vi finner indikasjon på en positiv effekt på elevenes prestasjoner etter den andre delen av reformen.

1 Introduction

School accountability is intended to reduce the principal-agent problem in education by providing *incentives* for teachers to boost student achievement and thereby school performance. The aim is that school accountability aligns the interests of the school district authority and the teachers, and induces a change in teacher behavior as the link between teacher effectiveness and school performance becomes more prevalent.

School accountability may also function as a *sorting* mechanism. School accountability might induce teacher mobility as student achievement is not directly attributable to teacher behavior. Many elements influencing student achievement are out of the teachers' control.[1] Making teachers accountable might therefore induce negative pressure and more risk on teachers, and hence trigger teacher mobility. Accountability might also crowd out teachers' intrinsic motivation.[2] In addition, performance-contracts in general lead to an increased administrative workload. This may induce disutility for some teachers. Teacher mobility might further affect the composition of teachers at the school level. If low-quality teachers move and are replaced by high-quality teachers, the sorting effect could be intentional, and it could increase overall teacher quality. However, high-quality teachers are not necessarily stayers.

In this paper, we study if school accountability is functioning as an incentive (for the incumbent workforce) or as a sorting mechanism (in terms of increased teacher turnover and changes in the composition of teachers at the school level). We exploit a management reform from 2003 that made schools *internally* accountable to the school district authority for student achievement, and the fact that a market-element was added in 2005, whereby information on school performance measured by conditional student achievement became public, thus made schools also *externally* accountable.

The literature on school accountability focuses mainly on channels through which school rankings can induce gaming responses from schools: teachers increase the use of special education placements (Jacob, 2005; Figlio and Getzler, 2002), substitute away from low-stakes subjects (Figlio, 2006), teach for the test (Jacob, 2005), cheat (Jacob and Levitt, 2003), and shift more attention to students in the middle of the achievement distribution (Neal and Schanzenbach, 2010) in order to inflate accountability scores. Feng et al. (2010) are one of few to study the effect of school accountability on teacher mobility.[3] They exploit

[1] See Kane and Stagier (2002) and Koretz (2002) who illustrate the pitfalls of imprecise school accountability measures.

[2] Extensive work by Deci and Ryan (e.g., 1985, 2000) indicates that too much control or distrust might negatively influence an individual's intrinsic motivation. Whereas school accountability is meant to give teachers more autonomy in the classroom, autonomy is coupled with measurement of school quality. School quality measures might be perceived as a signal of distrust. See Fehr and Falk (2002) concerning the psychology of incentives in general.

[3] There are some papers on school accountability and the mobility of school principals. E.g., Li (2012) finds that No Child Left Behind induced more able principals to move to schools less likely to face sanctions, thereby decreasing the average principal quality at schools serving disadvantage students. In addition, there are a few papers on school accountability

a change in Florida's school accountability system that exogenously shocked some schools to higher accountability scores and others to lower accountability scores. They find that teachers are more likely to leave schools that have been downwardly shocked and less likely to leave schools that have been upwardly shocked. Cooley and Traczynski (2013) study the dynamic effects of failing accountability and how sanctions inhibit school responses in North Carolina. They find that repeated failure means moving up sanctions levels, but no effect on teacher turnover.

When studying recruitment and turnover, economists have typically emphasized traditional pecuniary variables that affect labor demand and supply. Economists have to a lesser extent relied on organizational and social structures. However, there is a growing interest in such topics. Boyed et al. (2013) find that teachers prefer schools that are closer geographically to their home, are suburban, and have a smaller proportion of students in poverty, whereas schools prefer teachers with strong academic achievement and teachers living in proximity to the school. Boyed et al. (2011) find that teachers with better pre-service qualifications are more likely to apply for a transfer, while teachers whose students demonstrate higher achievements are less so.

Our paper extends the literature on school accountability and teacher mobility along two dimensions. First, we study teacher responses to two different accountability regimes. Hence, we are able to disentangle responses to two mechanisms (one internal and one external) and study how they trigger teacher behavior. Numerous studies have found that school accountability has a positive effect on student test scores (e.g., Rouse et al. 2013). Nonetheless, there is no real consensus (except perhaps for gaming) on the mechanism through which the impact of accountability takes place.[4] Second, in parallel to Lazear (2000) who study the effects of performance pay, we try to disentangle the sorting effect from the incentive effect. That is, we analyze how the distribution of observed teacher quality in terms of teachers' academic achievement[5] is affected by school accountability by studying teacher mobility and sorting. Furthermore, we try to see if a potential gain in student achievement is caused by the incumbent workforce or by a change in the composition of teachers.

---

and pupil sorting. E.g., Burgess et al. (2013) find indications of student sorting when school accountability was combined with school choice.

[4]Hanushek and Raymond (2004a) find that the effect of the publication of information dominates the incentive effect of sanctions and rewards. Hanushek and Raymond (2004b), on the other hand, find that just reporting results has minimal impact and that the force of accountability comes from attaching sanctions and rewards. Bishop et al. (2001) find that the "stick is more effective than the carrot". Harris and Herrington (2006) argue that the positive effects of accountability should mainly be attributed to the existence of exit exams. Rouse et al. (2013) show that improvement in student achievement can be attributed to changes in teaching practices.

[5]Several authors provide evidence of a positive influence of teachers' academic achievement on student achievement and hence that teachers' own grade from higher education can function as a proxy for teacher quality (e.g., Hanushek et al.2014; Hanushek and Rivkin, 2006; Clotfelter et al., 2006 and 2007). In addition, teachers' own academic achievement is a good indicator regarding teacher mobility and teachers' outside options as teachers own grades are salient for potential future employers.

There was no high-stake accountability testing in Oslo during the reform period. The school performance indicator was based on student grades from both teachers' evaluations and central exams. In Norway, student grades are only available in lower secondary education, so the nested accountability reform is likely to have affected primary and lower secondary teachers in Oslo differently, and thus induced a higher reform intensity for teachers in lower secondary education than for teachers in primary education. We first apply a differences-in-differences approach, comparing teachers outside the reform district to teachers comprised by the reform, before and after the reform was introduced. We then expand this to a difference-in-difference-in-difference (DDD)-approach, and compare the difference between treated teachers in lower secondary education and what we define as untreated teachers in primary education in the reform district to the difference between lower secondary and primary school teachers in school districts not affected by the reform. We use rich Norwegian data on public school teachers and students to study the causal effect of school accountability on teacher turnover, sorting, and student achievement.

Based on the DD-framework, the effects on teacher mobility are ambiguous. By exploiting the difference in reform intensity with a DDD-approach, we find a significant increase in teacher mobility in lower secondary education in Oslo after the internal part of the reform. The external part of the reform also triggers teacher turnover in lower secondary education, but not to a larger extent. Almost all non-stayers leave the teaching sector entirely. As regards the composition of teachers, high-ability teachers respond more strongly than low-ability teachers after the external part of the reform. Fortunately, after the external part of the reform, high-ability teachers are being replaced by high-ability teachers. We find a small, positive effect on student achievement after the external part of the reform. The channel of impact seems to be positive sorting (by teacher ability at the school level) rather than teacher incentives.

The paper proceeds as follows: Section 2 presents the institutional setting, reform details, and we discuss what can be expected of the nested accountability reform. Section 3 outlines the empirical strategy. Section 4 presents the data sources, defines important variables in the analysis, and presents some descriptive statistics. Section 5 outlines the empirical results concerning teacher mobility, sorting and student achievement, as well as heterogeneity analysis and robustness checks. Section 6 offers some concluding remarks.

2 Institutional setting and the nested accountability reform

2.1 The Norwegian educational system

Most schools in Norway are public.[6] Public schools have a common curriculum, the same number of teaching hours in each subject, and they are organized in

[6] More than 97 percent of the students are enrolled in public schools.

school districts. Whereas the governance structure can vary across districts, it is similar for primary and lower secondary school teachers within school districts. Teachers in Oslo, the reform district, are hired by the school they work at, but this is not the case for all schools in Norway.

The desirability of retaining and firing teachers may change as schools become more responsible for their performance. In Norway, the teacher labor market is strictly regulated, making it difficult to lay off teachers who have permanent positions. In addition, wage bargaining is centralized. There is little variation in wages across teaching jobs, and wages are difficult to use as a means of retaining teachers. In such an environment, mobility within the school sector will primarily be motivated by non-wage job attributes, as found by Falch and Strøm (2005).

In the litterature, alternative wages has been in focus when explaining out of sector mobility in the school sector (Dolton and van der Klaauw, 1995, 1999; Hoxby and Leigh, 2004). Chingos and West (2012) even find that teachers with high value added have higher earnings compared to other teachers who leave the teaching sector. In Norway, however, the wage structure is compressed and the returns to education are generally low, particular in the public sector (Barth and Moene, 2000). Inside the public sector, the teachers' external labor market is similar to the teacher labor market in terms of wages.

As regards the student body, schools take in students based on their catchment areas. The compulsory education track (i.e., primary and lower secondary education) starts at age 6 and continues until the age of 16. In contrast to many other countries, students in Norway are not graded before entering lower secondary education at the age of 13. In primary education, the evaluation of students is based on low-stakes tests only. There is no objective measure of school performance in primary education in the time period analyzed. Students in lower secondary education, on the other hand, are graded by their teachers in a total of ten subjects. In addition, students sit one central exit exam.[7] Grades from the last year of compulsory education are used to compete for study seats in upper secondary education.

2.2 The nested accountability reform

An emphasis on school performance was gradually implemented in Oslo from 2002 onwards. In 2002, there was a major reorganization in which school principals were granted substantive impact on school policies and hence assigned an important role in the process of generating educational success.

In 2003 (i.e., the internal part of the reform), school principals in Oslo became accountable to the school district authority for student achievement. Individual annual meetings were arranged between the authority and each school principal, at which school performance was discussed. Performance measurement was based on student achievements. Oslo had its own set of low-stake

[7] Students are randomly assigned to one examination among four subjects: Norwegian I and II, English and mathematics.

achievement tests. In lower secondary education, also student grades (both teachers' evaluations and central exams) were salient in this respect. In the case of low performance, school principals had to commit to changes in order to try to increase performance later on. Student grades are an accurate measure of performance, in addition to being easily interpretable to both teachers, principals, and the school district authority.

In 2005 (i.e., the external part of the reform), a market element was added to the accountability regime. First, a new adjusted school quality indicator was calculated for schools in lower secondary education, which aimed at indicating each school's contribution to student achievement. i.e., the value added. The indicator was based on mean grade points from both teachers' evaluations and central exams, and it was adjusted for individual student and parental characteristics (Hægeland et al., 2004). By adjusting the school quality indicator and including central exams, the scope for gaming by lower secondary education teachers, i.e., the possibility of inflating the accountability score artificially, was reduced. Second, the school quality indicator was publicly disclosed for the first and only time on November 18th. The aim was to inform parents and other stakeholders, and to further induce teachers to focus on school performance. At the time of the publication, both school principals and the public were told that there would not be any further public disclosures of school quality indicators. After the 2005 election, the new government strongly opposed public disclosure of school performance. Hence, the threat of further exposure for teachers in lower secondary education in Oslo was no longer imminent.

In 2006, a national reform implemented accountability mechanisms in all school districts, thereby aligning the system in Oslo with other school districts. In addition, a new performance measure with written assessments was implemented in primary schools in Oslo from 2006.

2.3 What to expect of the nested accountability reform?

Neither the internal part nor the external part of the reform are so-called high-stakes accountability regimes. Both are low-powered in the sense that no sanctions or rewards are attached. In addition, accountability tests are not included in the internal or the external part of the nested reform, as the performance indicator is based on regular student grades (both teachers' evaluations and central exams). This is in stark contrast to many of the accountability systems in the US and the UK where the ranking of schools are based on high-stakes testing and school accountability is associated with rewards, or the threats of firing teachers, replacement of principals, closure or reconstitution of schools, and allowing students to enroll elsewhere in the case of failure.

Even though the accountability reform is low-powered, the internal part of the reform did initiate a new way for school principals and teachers to govern and conduct schools.[8] School principals were made responsible for student

[8] The reform only targeted school principals and teachers. The accountability reform was not directed towards other stakeholders such as parents, school services, or health service staff.

achievement towards the school district authority, and teachers were the main channel through which they could fuel student achievements. School principals in Oslo needed therefore, in the wake of the nested school accountability reform, to inform and motivate teachers in parallel with delegating more responsibility and making teachers more accountable for student achievement. Furthermore, the incentive emanating from the external part of the reform is more high-powered than the incentive embedded in the internal part of the reform because student grades were made public. The external mechanism was added to the internal part of the reform, so the latter part of the reform could have provided enhanced incentives for teacher to increase student achievement.

Figlio and Loeb (2011) suggest that even absent of strong sanctions or rewards, accountability will affect both teaching environment and student achievement. Whereas sanctions and rewards are often regarded as necessary in order to change teacher behavior, school accountability can offer some interesting benefits. In fact, Dewatripont et al., (1999), by extending the one-task *career concerns*[9] model of Holmström (1982), find that total effort goes up when the number of tasks an individual has to perform decreases, indicating that the focus is effort-enhancing. The rationale is that accountability increases with the "clarity" of an organization's mission. In contrast, when an organization practices a "fuzzy mission", the market is uncertain about which mission an individual is actually pursuing. This theoretical result might indicate that teachers are able to develop stronger career concerns under the external part of the reform, suggesting that this part of the reform is more effective than the internal part in raising student performance. Even though both parts of the reform help to clarify the role of teachers and the mission of schools, only the latter part facilitates a market signal that might motivate teachers to put in more effort.

However, school accountability may not function as an incentive. Teachers might find the conditions under which they operate too inadequate in order to respond to the new regime: many elements influencing student achievement are out of their control. Accountability might also crowd out teachers' intrinsic motivation. The performance-contract between schools and the school district authority did also induce a higher administrative workload for teachers in Oslo compared to teachers in other parts of the country. Hence, the nested school accountability reform might have triggered teacher mobility.

It is hard to stipulate the effect on teacher sorting as both high and low quality teachers might move. Whereas high-quality teachers are considered to embody the necessary skills in order to respond to the new regime and thereby are more likely to stay than low ability teachers, they might find it hard to increase the overall school performance and might become demotivated. High-quality teachers (measured in terms of teachers' academic achievement) might

[9]Career concerns induce an individual to provide higher effort in the first period, when his or her innate abilities are neither known to themselves nor the market, in order to influence his or her prospects in the second period (Holmström, 1982). High effort and high performance in the first period normally correspond to a wage increase in the second period. For teachers in Norway, however, the wage is based on experience, so in our case career concerns might correspond to e.g., public recognition.

also have better outside options compared to low quality teachers as potential future employers value applicants with strong academic records. Hence, high-ability teachers might be more eager to move than low-ability teachers.

Lazear (2000) finds that, in the auto glass sector, switching from paying hourly wages to performance-based pay increased productivity by 44 %.[10] Half of the increase in productivity was attributed to incentives: the average worker worked harder after being paid on the basis of output. The other half was attributed to the ability to hire the most productive workers and a reduction in quits among the highest output workers. Regarding teachers, Dohmen and Falk (2010) find a diametrically opposed pattern. Their study suggests that introducing performance-based pay for teachers may crowd in teachers who are less trusting and more negatively reciprocal, at the cost of the current profile. As a consequence, the composition of teachers might negatively change and have an adverse effect on students' educational progress.

Concerning our nested school accountability reform, we cannot expect the same results (in terms of both productivity and sorting) as in the auto glass industry, given the more complex context of schooling and learning. However, the more team-based and more implicit (i.e., no sanctions or rewards) incentives inherent in the accountability regimes studied in this paper might be more suitable for teachers than the individualized and explicit incentive studied by Dohmen and Falk (*ibid.*).[11]

3 Empirical strategy

3.1 Teacher mobility

As a starting point, we first employ a Difference in Differences-framework to estimate the mean effect of the accountability reform on teacher mobility in the reform district. We compare the difference between teachers in the reform district to teachers outside the reform district before an after the implementation of the reform.

The following equation is estimated:

$$y_{it} = \beta_o + \beta_1 T_i + d_t + \gamma_1 (T d_{it}^T) + \beta_2 X_i + \varepsilon_{it} \quad (1)$$

[10] The effect was measured at the firm level and only concerned one firm. Probably, the effect would have been smaller if implemented at the industry level.

[11] Individual performance pay can be an adequate incentive for teachers (e.g., Lavy, 2009). Lavy (2015) even finds that teachers' pay for performance has positive long run effects on students' educational and labor market outcomes. In addition, Barley and Neal (2012) propose an incentive scheme for educators that rely on ordinal information contained in assessment results. They claim that such a scheme will reduce the gaming behavior of teachers as these schemes are more adequate than those relying on cardinal rankings. In general though, output-based incentives for teachers are often suggested to be low-powered in order to avoid gaming as high-powered explicit incentives are best used when output is well defined, the effort-performance relation is well understood, the production is uni-dimensional, and the outcome is easily measured (Dixit, 2002; Lazear, 2003).

The outcome variable y_{it} is a dummy for whether teacher i leaves the school in school year t or not. β_o is a constant. T_i is a dummy variable that equals one if teacher i is in the treated school district and zero if he/she belongs to a school district in the comparison group. d_t is a set of year dummies covering the period before, during and after the nested accountability reform (i.e., 2000-2006). d_t^T is a dummy variable equal to one if a reform year (i.e., 2004-2006) or zero otherwise. X_i is a vector of covariates that include gender, age, experience, controls for yearly local labor market conditions by educational background, a dummy for working in primary or lower secondary education, teacher education level, and dummies for having a teacher education at bachelor's and master's level. Age and experience are also included as a quadratic function. ε_{it} is a random error term clustered on school districts to safeguard against the possibility that the error term can be correlated within school districts.

Our parameter of interest in Equation 1 is γ_1. This parameter, in which the reform year dummy is interacted with treatment group status, measures the change in teacher turnover in the reform years relative to the years before the reform.

Quit decisions are made each year, so we also estimate a more general equation than Equation 1. Instead of an average reform effect, pooled over all reform years, Equation 2 contains year-specific effects. i.e., we replace the DD-parameter, γ_1, with a vector of year specific parameters, $\dot{\gamma_1}$:

$$y_{it} = \beta_o + \beta_1 T_i + d_t + \dot{\gamma_1}(Td_{it} \times d_{it}^T) + \beta_2 X_i + \varepsilon_{it} \tag{2}$$

To estimate what we can interpret as a causal effect on teacher mobility by the reform, we need to control for two kinds of potentially confounding trends: changes in teacher mobility across school districts (that have nothing to do with the nested reform) and changes in teacher mobility among teachers living in the policy-change school district (possibly due to other policies, or district-specific changes in the economy that affect all teachers in the treated school district). While the DD-approach do control for changes in mobility across school districts, it does not control for changes within the treated school district.

In fact, there are several concerns with our DD-approach. First, there are differences in the labor market in Oslo and the rest of the country. Large cities face a different pool of workers than the rest of the countries, and the labor market opportunities outside the teaching profession may be larger. If there are systematic difference in the treatment district and the comparison group, we may be concerned that they do not have a similar trend before the treatment, or that any common trend is purely a coincidence. Second, systematic differences in educational traits for primary and lower secondary school teachers could imply that common shocks in the labor market affect the two groups of teachers differently. Third, systematic differences in student composition in Oslo compared to the other school districts could also alter teachers' mobility response to common shocks.

Accountability was implemented in both primary and lower secondary education in Oslo. As student grades are salient to the school accountability

indicator and are not used in primary education, less pressure is induced on these teachers by the reform, and hence they are less likely to change workplace due to the implementation of school accountability.The reform intensity is thus assumed to be different in primary and lower secondary education.

That the same reform leads to different accountability pressure in primary and lower secondary schools is one of our main identifying assumptions. If our assumption does not hold, and primary school teachers are affected in the same way, any effects must be a result of other factors influencing the difference between primary and lower secondary school in Oslo compared to other school districts in the reform years. However, if primary and lower secondary school teachers are similarly affected by the reform, but react opposite, we will overestimate the effect of the reform.

As lower secondary education teachers are exposed to more pressure than what is the case for primary school teachers, we add a third difference to our emprirical strategy: We compare the difference between primary and lower secondary education teachers in the reform district to the same difference in the comparison district before and after the reform.

Any within-district differences are netted out by adding a third difference between primary school teachers and lower secondary school teachers. By comparing the difference in turnover response between primary and lower secondary school teachers in the reform district to the same difference outside the reform district, differences in the labor market are accounted for as primary and lower secondary education teachers are situated in the same labor market area. Systematic differences in educational traits are also accounted for by adding a third difference, as are any shocks that may affect primary and lower secondary education teachers differently. As the student body in the Oslo is the same for primary and lower secondary education, including the difference between primary school teachers and lower secondary school teachers controls for any differences in the student body.

In the DDD-framework, the following equation is estimated:

$$y_{it} = \beta_o + \beta_1 T_i + \beta_2 E_i + d_t + d_t E_i + d_t T_i + \gamma_1 (TE_i) + \gamma_2 (TEd_{it}^T) + \beta_3 X_i + \varepsilon_{it} \quad (3)$$

Our parameter of interest from Equation 3 is γ_2. This parameter, in which the reform year dummy is interacted with both treatment group status and lower secondary employment, measures the change in teacher turnover in the reform years (relative to the years before the reform) in the difference between turnover for primary school teachers and lower secondary teachers in- and outside the reform district. E_i is a dummy for working in lower secondary education that is interacted with year fixed effects and reform district.

We estimate the year specific effects of the reform by estimating Equation 4:

$$y_{it} = \beta_o + \beta_1 T_i + \beta_2 E_i + d_t + d_t E_i + d_t T_i + \gamma_1 (TE_i) + \dot{\gamma}_2 (TEd_{it} \times d_{it}^T) + \beta_3 X_i + \varepsilon_{it} \quad (4)$$

There might still be some concerns when applying a DDD-approach. If the reform led to strategic moving by parents, large changes in the moving patterns

could have led to changed student composition in the schools, which again could affect teacher mobility. Fiva and Kirkebøen (2011) find an increase in housing prices near high-quality schools in Oslo as a consequence of the 2005 reform, indicating strategic movings. They find however only a short-lived effect. A short lived effect could not have led to large changes in the schools' student composition. That our effect is driven by compositional changes due to the 2005 reform is therefore unlikely.

In general, there are indications of student composition influencing teachers' mobility decisions (e.g., Lankford, Loeb and Wyckoff, 2002; Hanushek, Kain, and Rivkin, 1999; Falch and Strøm, 2005). We therefore perform robustness tests to test whether changes in school characteristics over time drive our results by excluding small schools and schools with a high immigrant share. The labor market in the reform school district could also differ systematically from the labor market in other parts of Norway. If teachers in elementary schools and lower secondary schools are affected by the local labor market in the same way, it should not matter for our analysis. However, there are compositional factors, mainly that there are more lower secondary teachers with master's degrees than is the case for elementary schools (see Table A.1). This should be accounted for by controlling for the local unemployment rate within education level. The labor market for teachers in large cities could be systematically different from more rural areas. We test whether the results are sensitive to labor market regions in our robustness tests.

For our empirical strategy to be viable, there can be no other factors contributing to teacher mobility that affects only one combination of primary/lower secondary schools and Oslo/non-Oslo in the reform year. There were no other school reform or regime changes that applied to only one of these combinations in our estimation period.

Changes in the accountability regime could lead to sorting within the school sector, i.e., that teachers move to schools with higher performance or outside the treated school district. To find out whether changes in the turnover is a result of within-sector sorting, we also estimate Equation 4 with the outcome of leaving the teaching sector.

3.2 Heterogeneous treatment effects

On average, teachers may respond to increased accountability pressure by either choosing to stay put (and more or less align their behavior with the reform's intention) or to switch schools (in order to better tackle or avoid the new regime). However, under both the internal and the external part of the reform, the effect may vary across teacher sub-groups. That is, there might be gender differences, differences across age groups, education groups, among teachers with different levels of experience, and among teachers of different academic ability when it comes to responses to the nested reform.

Studying heterogeneous treatment effects could compliment the analysis as there are indications that groups of individuals respond differently to incentives (e.g., Leuven et al., 2010; Bettinger, 2010; Angrist et al., 2009; Angrist

and Lavy, 2009). In addition, they might have implications for teacher sorting (by ability). Heterogeneous treatment effects are estimated by running the regressions separately on sub-populations.

3.3 Teacher sorting

The overall effect on teacher composition depends not only on who leave, but also on the teachers replacing the ones who leave. To find out if there actually is a sorting effect of the accountability reform, we estimate the effect on the mean academic achievement of the stock of teachers in schools. We use a similar empirical approach as in Equation 4, but with mean academic achievement within the school as the outcome:

$$y_{it} = \beta_o + \beta_1 T_i + \beta_2 E_i + d_t + d_t E_i + d_t T_i + \gamma_1 (TE_i) + \gamma_2 (TEd_{it} \times d_{it}^T) + \beta_3 X_i + \varepsilon_{it} \quad (5)$$

The outcome y_{it} is the mean academic achievement of the employed teachers at time t for school i. All explanatory variables have the same interpretation as in Equation 4, with the exception of X_{it}, which now denotes a vector of control variables at the school level, including mean age, mean educational level, mean years experience, and male share. Our variable of interest is still γ_2.

Several authors provide evidence of a positive influence of teachers' academic achievement on student achievement and hence that teachers' own grade from higher education can function as a proxy for teacher quality (e.g., Hanushek et al.2014; Hanushek and Rivkin, 2006; Clotfelter et al., 2006 and 2007). Teachers with strong academic records are not always the same as those who actually boost student achievement, but when analyzing teacher mobility, teachers' own academic achievement is a good indicator for teachers' outside options as teachers own grades signals ability to future employers.

3.4 Student achievement

After studying teacher mobility and sorting, we investigate whether there are any effects on student achievement in order to disentangle incentives versus compositional impacts. As students are not graded before lower secondary education[12], we can not use the same DDD-approach as when studying teacher mobility. We run a DD-analysis by comparing student performance in lower secondary education in Oslo with the rest of the country before and after the implementation of school accountability:

$$y_{it} = \alpha_o + \alpha_1 T_i + d_t + \delta_1 \left(T_i d_t^T\right) + \alpha_2 X_{it} + \varepsilon_{it} \quad (6)$$

The outcome variable y_{it} measures the achievement of student i in year t. α_0 is a constant. T_i is a dummy variable equal to one if student i belongs to the treatment group, i.e., Oslo school district, and zero otherwise. d_t is a set of year dummies covering the period before, during and after the nested

[12] National test scores for primary education are only available from 2007.

accountability reform. d_t^T is a dummy variable equal to one if the year dummy is after a reform year (i.e., 2004, 2005, 2006). X_{it} is a vector of covariates (including dummy variables for age, gender, parental education and earnings, and immigrant category) . ε_{it} is a random error term clustered on school districts to safeguard against the possibility that the error term can be correlated within school districts. δ_1 captures the difference in change in student achievement in the reform period for the treated schools compared to the schools in the comparison group.

As we no longer include primary education in Oslo as a comparison group, it could be a concern that different teacher- and student characteristics shape the treatment and the comparison groups. Therefore, we also use a propensity score matching to find an appropriate comparison group as a robustness test.

4 Data and Descriptive Statistics

4.1 Teacher mobility and sorting

We use rich register data on public school teachers from Statistics Norway to study teacher turnover and sorting in Norway in the period between 2000 and 2006. Employment data on teachers include information on gender, age, education, experience (measured as years spent at the school), and appointment. We have employment data not only covering the reference week, but for every 4 weeks during the year. The data source also contains school identifiers and personal identification codes for each teacher. Since teacher mobility can be influenced by local labor market conditions, which may differ between the comparison group and the treatment group, we add yearly data on local unemployment by education level.

The sample for the mobility analysis is restricted to teachers eligible for permanent appointments. Eligible teachers who currently are in temporary positions are also included. Non-certified teachers are not allowed to keep permanent positions, and are subject to involuntary moves. We define eligible teachers as those who either have a teacher education, or has a teacher employment code. We also restrict the sample to those who work more than 50 percent of full-time. The pension age in Norway is 62 for most teachers, and we do not want to include those who leave the profession due to age. Therefore, only teachers between the ages of 20 and 60 are included in the sample for each year. Moreover, some schools in Norway are combined primary and lower secondary schools. We are not able to identify whether teachers at combined schools work in primary education or lower secondary education, and combined schools are therefore excluded. Schools that were closed down during the period are also excluded from our sample. No schools were closed down in the treatment area in our estimation period.

We add micro data on teachers' academic achievement. We construct an ability index from teachers' grades from higher education institutions (HEI), including all universities and university colleges in Norway. It is a strength that

HEI use external examiners from other institutions. A range of different grading scales is used for the grades included in the sample. We therefore normalize every grade within each grading scale.[13] We then calculate the mean grade achieved in HEI for each person, using all grades except for pass/fail.[14] We further adjust for institution-specific and field-specific effects, since grading practices vary across higher education institutions and study fields. Teachers with strong academic records are not always the same that actually boost student achievement since teacher quality is likely also to be defined by e.g., teacher effort, personal traits, and teaching practices. However, our ability index is a good measure of teachers' outside options (as teacher grades are important for potential new employers). When using teachers' academic achievement as a control variable, we divide the teacher population in two groups based on their academic achievement, high and low.

In total, the sample includes 64 306 teachers, 278 909 observations, for the years 2000-2006. Table A.1 in the appendix gives a descriptive overview of the main variables used in the analysis, for the total sample and for the treatment and comparison areas separately. As regards data on teachers' performance in higher education, we have information about at least one grade for 48 792 teachers in the sample.

The outcome variable in the mobility analysis is to leave the school, which is defined as not being registered as employed in the same school during the next calender year. Persons who have an end to their employment spell in a specific school during a year will not be registered as employed in the next year, and thus makes a transition. The exception is if they quit the job, but return to the school so that they are registered as employed in the school the next year. In that case, they will not be registered as making a transition by our definition. Most teacher mobility takes place during summer. For teachers who are employed at several schools at the same time, we chose what we define as the main employer (highest number of working hours and highest seniority). To leave the school thus includes both changing workplace within the treated area, changing jobs to other teaching jobs outside the treatment area, or leaving the sector entirely. Making schools accountable for student achievement to the school district authority in 2003, and publicly distributing new information about school quality in (November) 2005, are most likely to influence teacher turnover from 2004 and 2006, respectively.[15]

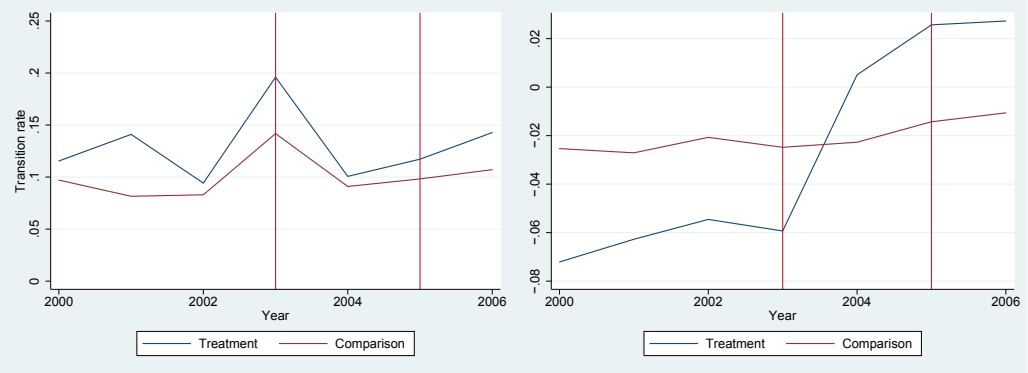
The critical assumption for both the DD- and the DDD-approach is that, in the absence of the nested reform, the difference between primary and secondary school teachers in the treatment and comparison groups follow a similar trend. The first part in Figure 1 shows no indication of a common pre-treatment trend in Oslo and the rest of the country. The second part in Figure 1, on the other hand, which shows the difference between primary and lower secondary

[13] A national grading system in HEI was first implemented in 2003.

[14] 60 percent of all grades included in the sample are obtained by teachers with exams in educational science.

[15] Teachers were not informed long before the implementation of each reform, i.e., teachers could not adjust their mobility responses ex ante, only ex post.

Figure 1: Transitions Rates in Oslo and the rest of the country and changes in the difference between Teacher Turnover in Lower Secondary and Primary Education in the Treatment and Comparison Districts



education teachers' mobility responses in the treatment and the control areas, confirms that there is indeed a common trend before the reform. In the pre-treatment period, teacher turnover in primary education is higher than teacher turnover in lower secondary education (see Figure A.1). That changes, however, in the treatment area during the treatment period. For the comparison group, there is no such shift as the common trend in mobility for primary and lower secondary school teachers also remains after 2003. The figures thus provide the first indication that the nested school accountability reform impacts primary- and lower secondary teachers in the reform district differently.

When estimating Equation 1 and Equation 2, there are negative effects on teacher mobility. However, the placebo test shows lack of a common pre-treatment trend, which is the main identifying assumption. Together, the failed placebo test and the problems discussed both here and in Section 3.1 indicate that the DD-framework is not reliable.

There is a spike in teacher mobility in 2003 for all teacher groups and in both areas, as seen in Figure A.1, indicating that there are national events affecting both the treated area and the comparison group. Such events could be surprisingly low achievement on the PISA test, an ex ante response to the introduction of national achievement tests[16], or business cycle conditions. Teacher unemployment reached a peak in 2003, which is coherent with the peak we find in our data. Tighter budget constraints at the school district level are the main reason for the high teacher unemployment in 2003. Such events should not influence our DDD-estimates, since it is unlikely that they would influence

[16] During our period, Norway participated in PISA in 2000, 2003, 2006 and 2009. Norway performed badly on the first PISA-test, and this is often referred to as the "PISA-shock". In addition, in 2004 national tests for students in 5th grade (primary education) and 9th grade (lower secondary education) were introduced, requiring annual testing of students in reading and mathematics.

the difference between teacher mobility in lower secondary schools and primary schools in Oslo differently from the same difference in other parts of the country.

Our main analysis ends in 2006 when the accountability regime in our comparison group changes. As a part of the robustness tests, long-term effects are also analyzed by adding data to 2008. With higher accountability intensity in primary education in Oslo, and the introduction of accountability regimes in the rest of the country, an increase in mobility in all parts of our comparison group could be expected after 2006.

The same data are used in the sorting analysis. We calculate mean academic achievement of teachers at the school level as our outcome variable. Positive sorting in terms of teacher ability implies that schools are able to attract and /or replace their high-ability teachers. A negative sorting effect implies that the reform schools are not able to attract and/or replace their high ability teachers. For this analysis, we have 11 337 observations from 2 378 schools.

4.2 Student achievement

We add data on student achievement as we attempt to disentangle sorting effects (increase in student achievement caused by a change in the composition of teachers) from incentive effects (increase in student achievement caused by the average teacher). In Norway, data on student achievement (i.e., teachers' evaluations and central exams) have been collected from 2002 onwards. Included in the data set are all grades of all students in the last year of lower secondary education.[17] In total, we have information on grades and social background variables for 278 223 students for the years 2002-2008.

No teacher value-added measures are calculated since we cannot link teachers and students. Thus we can not study the change in teacher composition at the classroom-level.

To measure the impact on student achievement, we construct an index including 10th-graders' average test score, based on grades obtained in Math, English and Norwegian, in addition to their central exam score. The test scores use a scale from 1-6 that we normalize with mean 0 and standard deviation 1 to facilitate interpretation of the results. This index corresponds to the unadjusted school quality indicator published for all schools in Oslo in November 2005.

5 Results

5.1 Teacher mobility

In Table 1, we see how lower secondary education teachers are affected by the reform, revealing a different picture. Column (1) in Table 1 reports estimated results based on Equation 3, whereas Columns (2) and (3) report variations of Equation 4. Control variables for teacher background are added in Column (3).

[17] Test scores on national tests for primary and lower secondary education are only available from 2007.

Table 1: The Effect of Accountability on Teacher Mobility, estimated by OLS

	(1)		(2)		(3)	
Oslo	0.026	(0.004)***	0.026	(0.004)***	0.017	(0.004)***
Lower Secondary, Oslo	-0.037	(0.002)***	-0.037	(0.002)***	-0.020	(0.002)***
Lower Secondary	-0.011	(0.005)**	-0.011	(0.005)**	-0.015	(0.004)***
Oslo*Lower Secondary*(2004-2005)	0.071	(0.004)***				
Oslo*Lower Secondary*2004			0.065	(0.004)***	0.052	(0.004)***
Oslo*Lower Secondary*2005			0.077	(0.005)***	0.066	(0.005)***
Oslo*Lower Secondary*2006	0.075	(0.005)***	0.075	(0.005)***	0.063	(0.005)***
R-squared	0.006		0.006		0.179	
Number of observations	278909		278909		278909	

Note: All specifications include a constant term, year dummies (ref. 2000), and the interaction terms $(T_i * d_t)$ and $(E_i * d_t)$. The third specification is used in all subsequent tables on teacher mobility, all estimated by OLS. Standard errors are clustered on school districts. */**/*** statistically significant at the 10/5/1 percent level.

Column (1) shows that the average treatment effect associated with the internal reform for lower secondary school teachers, is estimated to increase teacher mobility by 7 percentage points, from a pre-reform level of about 10 percent (see Table A.1). The same effect is found for the external part of the reform. Decomposing the average treatment effect of the 2003 reform (calculated for the period 2004-2005) into year-specific effects in Column (2) reveals the same picture. Publicly disclosing school performance seems therefore, on average, not to alter the response in terms of mobility for lower secondary school teachers relative to the response already emanating from the internal part of the reform.

Controlling for gender, age, experience, local unemployment (by education level), and teacher education (see Column 3) does slightly reduce the difference-in-differences-in-difference estimates. Adding teacher background variables does not change the overall picture of how school accountability triggers teacher turnover among lower secondary school teachers. As for the estimated coefficients of teacher background variables (not shown), they are all statistically significant, except for local unemployment conditions. The insignificant effect of local unemployment by educational level suggests that local labor market conditions are not a driving force inducing teacher mobility. On average, the control variables' contribution to explaining teacher mobility is low, except for experience, which is negatively related to teacher mobility.

5.2 Heterogeneous treatment effects

Although teachers in lower secondary education on average respond to the nested accountability reform by increasing their mobility during both accountability regimes, the effects may vary across subgroups. Table 2 shows the DDD-

estimates for different sub-samples; males versus females , young versus old (older than 35), experienced (more than 4 years at a particular school) versus less experienced, teacher education versus general education, and low-ability versus high-ability. These dimensions may be different in their response to the school accountability shocks, both with regards to alternative labor market prospects and in general responsiveness to incentives. Table A.2 indicates whether the differences between the subgroup pairs' DDD-estimates are statistically significant.[18]

We see that the teacher mobility responses of lower secondary school teachers are stronger for men than for women, although the difference between male and female teachers' responses is only significant in 2005. Young lower secondary school teachers change jobs to a significantly greater extent than older lower secondary school teachers after the external part of the reform. Hence, younger lower secondary school teachers respond more strongly to public exposure than their older counterparts. As regards experience, teachers with little experience (less than 4 years at a particular school) move to a much greater extent than teachers with more experience, although the difference is not significant in 2005.[19]

There are two main education tracks for becoming a teacher in Norway: a track with an educational focus, i.e. teacher education, and a more general track for those who become teachers after first taking a degree in a subject area of interest. The former education track implies that the decision to go into teaching is taken before starting higher education, whereas, in the latter track, this decision may be postponed as they first complete a program not specifically oriented towards teaching, and then supplement it with one year of specialized teacher training. Moreover, teachers with a teacher education might have fewer outside options than teachers with a a more general education. We find that lower secondary school teachers with teacher education move to a significantly higher extent than lower secondary school teachers with a more general education in 2004.

Lower secondary school teachers with strong academic records react stronger than those with academic achievement below mean in 2004 and 2006.[20] In 2005 there is no significant difference in the mobility response for the two groups. When we see a stronger reaction in the high-ability group, an important reason may be that those with higher academic achievement also have better labor market prospects as academic achievement serves as a signal of ability. The dif-

[18] We test the linear combination of two estimates.

[19] Those with long experience count for the majority of the teachers in our sample and display a turnover effect close to our average effect. If teachers are not granted a permanent position when first appointed, they will automatically get a permanent position after staying four years in a particular school. That we re-find our average reform effect on a sub-sample of teachers with only permanent positions indicates that our mobility effect is not driven by selective dismissals by school principals.

[20] We do not have data on teacher academic achievement for the whole sample, see Section 4.1. Heterogeneous effects concerning high and low ability teachers are therefore based on a smaller sample. We re-run the main specification (specification 3, Table 1) on the smaller ability index sample. This does not alter our average mobility effect (not shown).

ference in the 2006-effect is substantial, especially considering that the baseline mobility is very similar both for high-ability and low-ability teachers. Other studies have also found that teachers with high test scores more often leave the teaching profession than teachers with lower scores do, although not as a consequence of school accountability (Murnane and Olsen, 1990; Henke et al., 2000; Podgusrsky et al., 2004; Boyd et al., 2010).[21]

Table 2: Heterogeneous Treatment Effects, Teacher Mobility

	male		female		young		old	
Oslo*Lower Secondary								
*2004	0.064	(0.005)***	0.048	(0.005)***	0.052	(0.007)***	0.050	(0.004)***
*2005	0.101	(0.007)***	0.050	(0.005)***	0.055	(0.008)***	0.069	(0.005)***
*2006	0.064	(0.007)***	0.053	(0.005)***	0.078	(0.008)***	0.046	(0.005)***
R-squared	0.220		0.204		0.234		0.160	
Number of observations	79372		199537		90278		188631	
Baseline mobility	10.83		11.52		15.75		7.88	
	long experience		short experience		Teacher Education		general education	
Oslo*Lower Secondary								
*2004	0.039	(0.004)***	0.114	(0.016)***	0.043	(0.004)***	0.003	(0.012)
*2005	0.058	(0.004)***	0.087	(0.016)***	0.050	(0.005)***	0.034	(0.012)***
*2006	0.042	(0.004)***	0.127	(0.017)***	0.037	(0.005)***	0.040	(0.013)***
R-squared	0.060		0.054		0.202		0.272	
Number of observations	232036		46873		250316		28593	
Baseline mobility	2.42		24.12		12.63		9.40	
	high academic achievement		low academic achievement					
Oslo*Lower Secondary								
*2004	0.074	0.006	0.039	0.006				
*2005	0.061	0.007	0.057	0.006				
*2006	0.097	0.007	0.032	0.006				
R-squared	0.197		0.235					
Number of observations	106 530		99 306					
Baseline mobility	11.04		12.72					
Note: see Table 1								

[21] We also analyze the effect of new school performance information on teacher mobility. We compare the ranking of schools based on the adjusted and the non-adjusted performance indicator related to the external part of the reform. We find no average effect of positive- or negative information shocks on teacher mobility, but in schools receiving a negative information shock, low-ability teachers experience reduced mobility, while for high-ability teachers in the same schools, the mobility increases.

5.2.1 Out of sector mobility

So far, we have studied whether lower secondary school teachers change workplaces or not. An alternative outcome is the extent to which lower secondary school teachers leave the sector entirely. That is, do lower secondary school teachers move into other teaching jobs (and strategically move in or out of the treatment group) or do they leave the school sector entirely? Table A.3 shows that most of those who change jobs actually leave the teaching profession.[22] The same mobility effects (results not shown) are found for this alternative outcome variable (out of sector mobility) as for the main outcome variable (change in workplace). In contrast to previous studies, we do not find that those who leave the teaching profession often leave employment altogether (Stinebrickner, 2002; Fritjers et al., 2004). Few go to better paid jobs, which is coherent with non-wage attributes driving teacher mobility as discussed in Section 2.1.

5.3 Robustness checks

5.3.1 Placebo and alternative comparison groups

We conduct several robustness checks to investigate the sensitivity of our findings. First, we perform a placebo test. Based on Equation 4, we test for plausible reform effects in the years before the nested accountability reform. Reform effects should not be found before the implementation of the nested accountability reform. Table 3 shows the year-specific effects for lower secondary education in Oslo before, during and after the implementation of the nested accountability reform. The DDD-estimates are indeed insignificant in the pre-treatment years 2001, 2002 and 2003. The difference between lower secondary education teachers and primary school teachers thus have a common trend before the implementation of the reform in Oslo and the rest of the country.

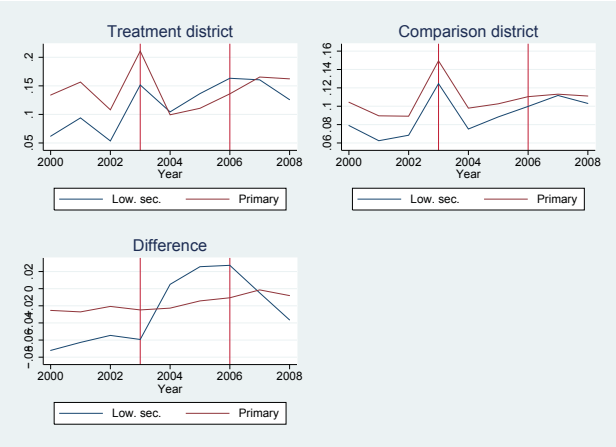
Table 3: Placebo Testing, Teacher Mobility

	Specification (3)	
Treatment Effect (ref. 2000)		
- 2001	0.003	(0.006)
- 2002	0.003	(0.006)
- 2003	-0.004	(0.006)
- 2004	0.052	(0.006)***
- 2005	0.063	(0.007)***
- 2006	0.059	(0.006)***
R-squared	0.207	
Number of observations	278909	

Note: see Table 1

[22] There are too few observations to study other transitions, such as mobility in or out of treatment.

Figure 2: Long term transition rates



We investigate further whether the results are sensitive to the choice of comparison group. We first exclude small schools (less than 20 persons in full-time positions per school) as a robustness check as there are few small schools in the treatment group. This does not change our DDD-estimates. It might still be a concern that the labor market for teachers in lower secondary education is different than for teachers in primary education, and that there are differences in labor market conditions in Oslo compared to the rest of the country. We therefore change the comparison group to first only include school districts around Oslo (which are part of the same labor market region), and then to only include the main cities in Norway (which might have similar and, on average, better pools of applicants). None of these changes in the comparison group influence our DDD-estimates (results not shown).

5.3.2 Long-term effects

No long-term effects (for the years 2007-2008) of the nested accountability reform are found in lower secondary education in Oslo. The effect fades out in 2007 and is non-existent in 2008 (results not shown). By studying Figure 2, we see that the transition rate for teachers in lower secondary education in the reform district decreases after 2006, while it rises for teachers in primary education in the reform district. Written performance assessments were introduced for primary schools in 2006, thus increasing accountability intensity for these teachers. An increase in transition rates is also observed in the comparison area post 2006, which could be expected as a consequence of the 2006 national school reform. From 2006, there is an alignment of accountability systems across the country, which is coherent with what we see in the data. In this case, the lack of long-term effects therefore strengthens our argument that accountability does in fact increase teacher mobility.

Table 4: The Effect of Accountability on Teacher Sorting, estimated by OLS

	Rest of the country as comparison group		Large cities as comparison group	
Oslo	-0.030	0.006***	-0.021	0.008
Lower Secondary, Oslo	-0.005	0.009	-0.026	0.016
Lower Secondary	-0.008	0.013	0.018	0.018
Oslo*Lower Secondary*2004	-0.005	0.010	-0.002	0.012
Oslo*Lower Secondary*2005	0.017	0.010	0.017	0.016
Oslo*Lower Secondary*2006	0.023	0.012**	0.018	0.018
R-squared		0.018		0.039
Number of observations		11 337		2 855

Note: See Table 1

5.4 Teacher sorting

Which types of teachers sort into schools under school accountability? The previous analysis were concerned with the outflow of teachers, and notably the teacher turnover among high-ability teachers. We now also study joining teachers' ability by focusing on the changes in the ability distribution of the stock of teachers at the school level. By estimating Equation 5, i.e., measuring the effect on teacher academic achievement at the school level of being in a treated lower secondary school in a reform year, we disclose if there is any sorting by teacher ability at the school level following the reform.

The mean teacher ability (as measured by teacher academic achievement) increases in the reform schools in 2005 (not statistically significant) and 2006 (statistically significant), as seen in Table 4. This means that, even though school accountability does not encourage the right pattern of retention as seen in Section 5.2, schools in Oslo are able to attract high-ability teachers. The positive effect implies that in terms of ability, the reform gives rise to positive sorting in Oslo relative to the rest of the country.

Large cities face a different pool of potential applicants for available teacher positions than the rest of the country. As sorting may be different in the large cities than in the rest of the country, we repeat the analysis only with the main cities as comparison group. We find the same pattern, although precision is lost, and the results are no longer significant. When performing placebo tests, no significant effects are found in either cases for the pre-reform years (results not shown).

Even though Oslo could be able to attract applicants with strong academic background, this may not be the reality for all parts of the country, and other school districts may face more severe challenges with negative sorting. Also low-performing schools in Oslo are able to attract high-quality teachers. Again, this may not be the case for other parts of the country. Other studies find adverse effects on teacher turnover in low-performing schools (e.g., Clotfelter et al., 2004) and adverse effect on school principal mobility in low-performing schools (e.g.,

Table 5: The effect of Accountability on Student Achievement, estimated by OLS

	Specification (1)		Specification (2)	
Oslo	0.106	(0.016)***	0.022	(0.009)**
Oslo*2004	-0.031	(0.007)***	-0.039	(0.009)***
Oslo*2005	-0.004	(0.007)	-0.007	(0.006)
Oslo*2006	0.035	(0.007)***	0.031	(0.007)***
R-squared	0.001		0.226	
Number of observations	278 228		278 223	

Note: All specifications include a constant term and year dummies (ref. 2008). Specifications are estimated by OLS. Standard errors are clustered on school districts. */**/*** statistically significant at the 10/5/1 percent level.

Li,2012). Dizon-Ross (2014), on the other hand, provides evidence concerning accountability in disadvantaged schools that is more aligned with ours. She finds that a lower accountability grade among schools at the bottom end of the school grade distribution decreases teacher turnover among high-quality teachers and increases joining teacher's quality, whereas a lower accountability grade among schools at the top end of the school grade distribution has no turnover effect, but decreases joining teachers' quality.

5.5 Student achievement

We study next if there are any effects on student achievement and whether the net policy impact consists of compositional versus incentive effects. Table 5 shows the difference-in-difference estimates based on Equation 6. Specification (1) includes no control variables, whereas specification (2) includes both teacher and student characteristics.

We also perform a placebo test to check if there is a common trend before the reform, as seen in Table A.4. We find a significantly positive coefficient on the effect in 2003, before the reform is in place. Also when using a difference-in-differences approach for teacher mobility, we were not able to confirm a common trend before treatment. With Oslo having a different set of students and labor market opportunities than the rest of the country, such a finding is not surprising.

There are only two years of observations pre-reform, and the findings on student achievement is therefore difficult to interpret. If more data were available, a more thorough analysis on student achievement could been done.

In Table 5, there is a negative effect of being in a lower secondary school in the reform district in 2004. In 2006, after the external part of the reform, it shifts to a small, but significantly positive effect. The effect amounts to about 3 percent of a standard deviation.

The results on student achievement are to some extent coherent with the

findings on teacher sorting, with indications of a positive effect after the external part of the reform. These results could suggest that, on average, publicly disclosing school rankings have a positive impact on school performance, and that positive teacher sorting has an impact on student acievement.[23] For student achievement, however, care must be taken when interpreting the results, as they are not very robust, i.e., failed placebo test in 2003.

Concerning control variables (not shown), teachers with high academic achievement contribute on average positively to student achievement. Moreover, the coefficient related to the dummy variable for whether teachers belong to the inflow-category or not (i.e., the outflow/ not switching workplace-category) is not significant. This might indicate that, on average, teacher turnover (given our sorting pattern) has no effect on student performance.

Oslo has a higher share of immigrant students, more dispersed social background of the student body, and more teachers with a master's degree compared to the rest of the country. As a robustness test, we construct an appropriate comparison group based on propensity score matching; matched on characteristics for parental education, migration characteristics (migration age and migration area), and teacher characteristics (gender and education). Using such a comparison group does not alter our results.[24]

We find that the strongest impact on student achievement relates to an increase in the grades awarded to students by their teachers in English.[25]

6 Concluding remarks

It is essential to understand teacher mobility and sorting if we want to design adequate incentives for teachers and comprehend school performance. In this paper, we have studied two different accountability regimes, one internal and one external, and we have evaluated their effects on teacher mobility, sorting, and student achievement.

We find a significant positive effect on teacher mobility in the years after

[23] Carnoy and Loeb (2002) find a positive and significant relationship between the strength of states' accountability systems and achievement gains. Hanushek and Raymond (2004a) find a positive impact of school accountability as a result of the publication of information on school performance. These results are somewhat congruent with ours as the external reform seems to be more effective than the internal reform. These authors, however, do not try to disentangle sorting and incentive effects and attribute the (net) policy impact to teacher incentives.

[24] We match on school level, using nearest neighbor. The same method of matching is also used in the mobility analysis, which do not alter the estimates (results not shown).

[25] Students' teacher-awarded grades are at least partially within the control of teachers and are therefore more manipulable than test scores. We find, however, no indication of teachers in Oslo inflating students' teacher-awarded grades as they do not seem to increase more relative to scores on central exams. The design of the performance indicator, no high-stake accountability tests, and excluding sanctions and rewards may have facilitated a non-gaming behavior of teachers. Whereas students' teacher-awarded grades are more manipulable than test scores, they might on the other hand be harder to influence without gaming, e.g., teaching to the test is not possible. Notably, this might explain the low positive impact of school accountability on student achievement that we find in Oslo.

the internal part of the reform. The external part of the reform do not trigger teacher turnover to a higher extent than the internal reform, except for younger teachers. Moreover, the effects of the nested accountability reform are stronger for male teachers, and the majority of teachers who change jobs leave the public school sector entirely.

As regards the impact on teacher sorting, high-ability teachers move more than low-ability teachers after the external part of the reform. Fortunately for the reformed school district, the high-ability teachers who left after the external part of the reform were largely replaced by high-ability teachers, indicating that even though there were adverse effects on teacher turnover, schools in Oslo did benefit from an inflow of equally able teachers (as measured in terms of teacher grades from higher education).

We find indication of a small, positive effect on student achievement after the external part of the reform. Teacher sorting is likely to have contributed to such an effect.

With suitable data, we hope in the future to also study the sorting and the incentive effects at the classroom level, by linking teachers and students. Further studies of the effect of a particular (math) teacher on student achievement (in math) by using both the value added measure and controlling for teachers' own grades (in math) are also necessary in order to see how strongly the two measures of teacher quality (value added and teachers' own grades from higher education) are correlated.

References

[1] Angrist, J. and Lavy, V. (2009). The Effects of High Stakes School Achievement Awards: Evidence from a Randomized Trial. *American Economic Review* 99, 301–331.

[2] Angrist, J., Lang, D., and Oreopoulos, P. (2009). Incentives and Services for College Achievement: Evidence from a Randomized Trial. *American Economic Journal: Applied Economics* 1,1,1-28.

[3] Barlevy G. and Neal D. (2012). Pay for Percentile. *American Economic Review* 102, 1805-1831.

[4] Barth, E. and Moene, K.O. (2000). Er lønnforskjellene for små? NOU No. 21, Oslo, Ministry of Finance.

[5] Bettinger, E.P. (2012). Paying to learn: The effect of financial incentives on elementary school test scores. *Review of Economics and Statistics* 94, 686-698.

[6] Bishop, J. H., Mane, F., Bishop M., and Moriarty, J. (2001). The Role of End-of-Course Exams and Minimal Competency Exams in Standards-Based Reforms. In Brookings Papers in Education Policy, edited by Diane Ravitch, 267-329.

[7] Boyd, D., Lankford, H., Loeb, S., and Wyckoff. J. (2013). Matching Public School Teachers to jobs: Disentangeling the Preferences of Teachers and Employers. *Journal of Labor Economics* 31,1, 83-117.

[8] Boyd, D., Lankford,H., Loeb, S., Ronfeldt, M. and Wyckoff, J. (2011). The Role of Teacher Quality in Retention and Hiring: Using application-to-transfer to uncover preferences of teachers and schools. *Journal of Policy Analysis and Management* 30,1, 88–110.

[9] Burgess, S., Wilson, D., and Worth J. (2013). A Natural Experiment in School Accountability: the impact of school performance information on pupil progress and sorting. *Journal of Public Economics* 106, 57-67.

[10] Carnoy, M. and Loeb, S. (2002). Does External Accountability Affect Student Outcomes? A Cross-State Analysis. *Educational Evaluation and Policy Analysis* 24, 4, 305-331.

[11] Chingos, M. and West, M. (2012). Do More Effective Teachers Earn More Outside the Classroom? *Education Finance and Policy* 7,1, 8-43.

[12] Clotfelter, C.T., Ladd, H.F., Vigdor, J.L., and Diaz, R.A. (2004). Do School Accountability Systems Make it more difficult for Low-Performing Schools to Attract and Retain High-Quality Teachers? *Journal of Policy Analysis and Management* 23, 2, 251-271.

[13] Clotfelter, C.T., Ladd, H.F., and Vigdor, J.L. (2006). Teacher-Student Matching and the Assessment of Teacher Effectiveness. *Journal of Human Resources* 41, 4, 778–820.

[14] Clotfelter, C.T., Ladd, H.F., and Vigdor, J.L. (2007). How and Why Do Teacher Credentials Matter for Student Achievement? NBER Working Paper 12828.

[15] Cooley, J. and Tranczynski, J. (2013). Spare the Rod? The Dynamic effects of failing accountability on schools. Working Paper.

[16] Deci, E. L. and Ryan, R. M. (1985). Intrinsic motivation and self-determination in human behaviour. New York: Plenum.

[17] Dewatripont M., Jewitt, I., and Tirol, J. (1999). The Economics of Career Concerns, Part II: Applications to Missions and Accountability of Government Agencies. *Review of Economic Studies* 66, 199-217.

[18] Dizon-Ross, R. (2014). How Do School Accountability Reforms Affect Teachers? Evidence from New York City. Workingpaper.

[19] Dixit, A. (2002). Incentives and Organizations in the Public Sector: An Interpretative Review. *Journal of Human Resources* 37, 4, 696-727.

[20] Dohmen, T. and Falk, A. (2010). You get what you pay for: Incentives and Selection in the Education System. *Economic Journal* 120, 256-271.

[21] Dolton, P. and van der Klaauw, W. (1995). Leaving teaching in the UK: A duration analysis. *Economic Journal* 105, 429, 431-44.

[22] Dolton, P. and van der Klaauw, W. (1999). The turnover of teachers: A competing risks explanation. *Review of Economics and Statistics* 81 ,3, 543-52.

[23] Falch, T. and Strøm, B. (2005). Teacher turnover and non-pecuniary factors. *Economics of Education Review* 24, 611-631.

[24] Feng, L., Figlio, D., and Sass, T. (2010). School Accountability and Teacher Mobility. NBER Working Paper 16070.

[25] Fehr, E. and Falk, A. (2002). Psychological foundations of incentives. *European Economic Review* 46, 687-724.

[26] Fiva, J.H. and Kirkebøen, L.J. (2011). Information Shocks and the Dynamics of the Housing Market. *Scandinavian Journal of Economics* 113, 3, 525-552.

[27] Figlio, D. (2006). Testing, Crime and Punishment. *Journal of Public Economics* 90, 4-5, 837-851.

[28] Figlio, D. and Getzler, L. S. (2002). Accountability, Ability and Disability: Gaming the System. NBER Working Papers 9307.

[29] Figlio, D. and Loeb, S. (2011). School Accountability. Handbook in Economics of Education, Volume 3, Amsterdam: North Holland, 383-421.

[30] Fritjers P., Shields, M., and Price, S.W. (2004). To teach or not to teach? Panel data evidence on the quitting decision. IZA DP No. 1164.

[31] Grogan E. and Youngs, P. (2011). Fitting in: Person-Organization, Person-Job, and Person-group Fit as Drivers of Teacher Mobility. Working Paper No. 21, The Education Policy Center at Michigan State University.

[32] Hanushek, E., Piopiunik M., and Wiederhold, S. (2014). The Value of Smarter Teachers: International Evidence on Teacher Cognitive Skills and Student Performance. NBER Working paper 20727.

[33] Hanushek, E., Kain, J.F, and Rivkin S. (1999). Do Higher Salaries Buy Better Teachers? NBER Working paper 7082.

[34] Hanushek, E. and Raymond, M. (2004a). The Effect of School Accountability Systems on the Level and Distribution of Student Achievement. *Journal of the European Economic Association* 2, 2-3, 406-415.

[35] Hanushek E. and Raymond, M. (2004b). Does School Accountability Lead to Improved Student Performance? *Journal of Policy Analysis and Management* 24, 2, 297-327.

[36] Hanushek E. and Rivkin, S.G. (2006). Teacher Quality. Handbook of the Economics of Education, Volume 2, Amsterdam: North Holland, 1052-1078.

[37] Harri, D. and Herrington, C. (2006). La régulation par les résultats contribue-t-elle à l'amélioration des écoles? Améliorer l'école (15), D. Meuret et G. Chapelle (eds.), Presses universitaires de France, 204-213.

[38] Henke, R. R., Chen, X., Geis, S., and Knepper P. (2000). Progress through the teacher pipeline: 1992-93 college graduation and elementary/secondary school teaching as of 1997. NCES 152, Washington DC, US Department of Education.

[39] Holmstrom, B. (1982). Managerial Incentives - a Dynamic Perspective. in Essays in Honor of Lars Wahlbeck, Helsinki: Swedish School of Economic.

[40] Hoxby, C.M. and Leigh, A. (2004). Pulled Away or Pushed Out? Explaining the Decline of Teacher Aptitude in the United States. *American Economic Review* 94, 2, 236-240.

[41] Hægeland, T., Kirkebøen, L., Raaum, O., and Salvanes K.G. (2004). Marks across Lower Secondary Schools in Norway: What can be explained by the Composition of Pupils and School Resources? Report No. 11, Statistics Norway.

[42] Jacob, B. and Levitt, S. (2003). Rotten apples: An investigation of the prevalence and predictors of teacher cheating. *Quarterly Journal of Economics* 118, 843-877.

[43] Jacob, B. (2005). Accountability, Incentives and Behavior: Evidence from School Reform in Chicago. *Journal of Public Economics* 89, 5-6, 761-796.

[44] Kane, T. J. and Staiger, D. O. (2002). The Promise and Pitfalls of Using Imprecise School Accountability Measures. *Journal of Economic Perspectives* 16, 91-114.

[45] Koretz, D. (2002). Limitations in the Use of Achievement Tests as Measures of Educator' Productivity. *Journal of human resources* 37, 752-777.

[46] Lankford H., Loeb, S., and Wyckoff, J. (2002). Teacher Sorting and the Plight of Urban Schools: A Descriptive Analysis. *Educational Evaluation and Policy Analysis* 24, 1, 37-62.

[47] Lavy, V. (2009). Performance Pay and Teachers' Effort, Productivity, and Grading Ethics. *American Economic Review* 99, 5, 1979-2011.

[48] Lavy, V. (2015). Teachers' Pay for Performance in the Long Run: Effects on Students' Educational and Labour Market Outcomes in Adulthood. NBER Workingpaper.

[49] Lazear, E.P. (2000). Performance Pay and Productivity. *American Economic Review* 90, 5, 1346-1361.

[50] Lazear, E.P. (2003). Teacher Incentives. *Swedish Economic Policy Review* 10, 179-214.

[51] Leuven, E., Oosterbeek, H., and van der Klaauw, D. (2010). The effect of financial rewards on students' achievement: Evidence from a randomized experiment. *Journal of the European Economic Association* 8, 6, 1243–1265.

[52] Li, D. (2012). School Accountability and Principal Mobility: How No Child Left Behind Affects the Allocation of School Leaders. Working Paper.

[53] Murnane, R.J. and Olsen R.J. (1990). The effects of salaries and opportunity cost on length of stay in teaching. Evidence from North Carolina. *Journal of Human Resource* 25, 106-124.

[54] Neal, D. and Schanzenbach D.W. (2010). Left behind by design: proficiency counts and test-based accountability. *Review of Economics and Statistics* 92, 263-283.

[55] Podgursky, M., Monroe,R., and Watson, D. (2004). The academic quality of public school teachers: an analysis of entry and exit behavior *Economics of Education Review* 23, 507-518.

[56] Rouse, C.E, Hannaway J., Goldhaber D., and Figlio D. (2013). Feeling the Florida heat? How low-performing schools respond to voucher and accountability pressure. *American Economic Journal: Economic Policy* 5, 251-281.

[57] Ryan, R. M. and Deci, E. L. (2000). Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American Psychologist* 55, 68-78.

[58] Stinebrickner T. (2002). An analysis of occupational change and departures from the labour force. *Journal of Human Resources* 37, 192-216.

Appendix

Table A.1: Descriptive Statistics for the estimated sample (fractions unless otherwise noted)

	SAMPLE	OSLO		OTHER SCHOOL DISTRICTS	
		LOWER SECONDARY	PRIMARY	LOWER SECONDARY	PRIMARY
OUTCOME VARIABLE					
Transition (2000-2006)	10.24	10.99	13.71	8.60	10.61
Transition (2000-2003)	8.95	7.02	13.32	6.97	9.41
EXPLANATORY VARIABLES					
Male	28.46	38.41	18.92	42.36	23.01
Age (average)	42.78	42.51	40.55	43.32	42.76
Experience (average)	10.52	9.53	9.20	11.10	10.42
Unemployment (average)	0.014	0.020	0.020	0.013	0.013
Education at Master's level	10.25	39.32	6.78	21.36	4.85
Teacher Education at Bachelor's level	87.92	57.97	89.87	76.59	93.58
Teacher Education at Master's level	1.83	2.72	3.35	2.05	1.57
Number of observations	278909	5598	16598	77 601	179 112
Number of teachers	64306	1426	4413	19390	43270

Note: The number of teachers in the different subgroups does not add up to the total number in the sample due to mobility across groups.

Figure A.1: Trends before, during and after the nested reform: teacher mobility separately for primary- and lower secondary teachers

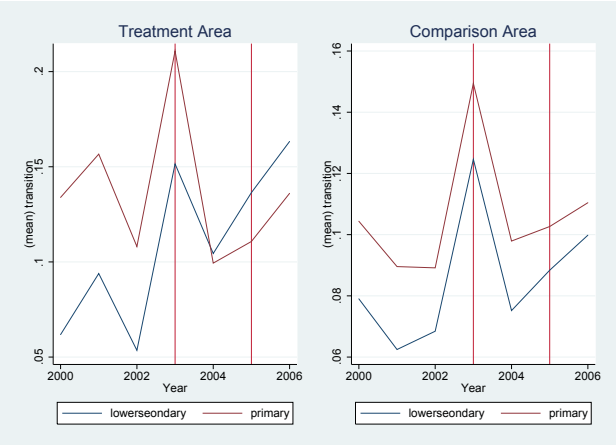


Table A.2: Testing for Statistically Significant Differences

Differences		
Male - Female		
2004	0.017	(0.007)
2005	0.051	(0.007)***
2006	0.010	(0.007)
Young - Old		
2004	0.002	(0.008)
2005	-0.013	(0.008)
2006	0.032	(0.008)***
Short - Long Experience		
2004	0.073	(0.016)***
2005	0.030	(0.016)
2006	0.087	(0.017)***
Teacher education -General education		
2004	0.040	(0.013)***
2005	0.017	(0.013)
2006	-0.002	(0.015)
Low - High academic achievement:		
2004	0.021	(0.0073)***
2005	-0.010	(0.0085)
2006	0.050	(0.0064)***

Table A.3: Types of Transitions, Teacher Mobility

	Transitions	Percent
Stay in the same school	250 349	89.76
New school, same school district	5 030	1.80
New school, new school district	4 070	1.46
Leave school sector	19 462	6.98

Table A.4: Placebo Testing, Student Achievement

	Specification (2)	
Treatment Effect (ref. 2008)		
2002	-0.014	(0.009)
2003	0.042	(0.010)***
2004	-0.030	(0.010)***
2005	0.002	(0.009)
2006	0.040	(0.008)***
2007	0.006	(0.007)
R-squared	0.226	
Number of observations	278 223	

Note: see Table 5

Statistics Norway

Postal address:
PO Box 8131 Dept
NO-0033 Oslo

Office address:
Akersveien 26, Oslo
Oterveien 23, Kongsvinger

E-mail: ssb@ssb.no
Internet: www.ssb.no
Telephone: + 47 62 88 50 00

ISSN: 1892-753X

Statistisk sentralbyrå
Statistics Norway

Design: Siri Boquist