Wages, Teacher Qualifications and Student Achievement

Hege Marie Gjefsen

Preliminary draft, do not cite

Abstract

I exploit a benefit program in disadvantaged schools in the city of Oslo to estimate the effect of increased wages in primary and secondary education. The program increased wages with up to 5%, in addition to reduced workload for less experienced teachers. Using a difference-in-differences approach, I study whether the program changed characteristics of the newly hired teachers in terms of educational background and academic achievement, and whether it improved student achievement. I find positive effects on the probability of hiring teachers with master degrees, including master degrees in science. Results with different comparison groups indicate that this is not a sorting effect, but rather an effect on the extensive margin. There are also indications of improved student achievement in the reform schools.

1 Introduction

Difficulties with recruiting qualified teachers in disadvantaged schools suggest a negative teacher sorting, which impedes the ideal of equal educational opportunities. Teacher workload is typically considered higher in disadvantaged schools. Hence, teachers will sort out of schools with the highest workload, everything else equal. To overcome teacher sorting out of disadvantaged schools, increased wages and/or better working conditions could give disadvantaged schools a competive advantage when it comes to teacher recruitment.

To counteract difficulties with recruiting qualified teachers, disadvantaged schools in the city of Oslo have been granted a benefit program, of which increased wages are the most important component. I exploit the program to estimate effects of higher wages on teacher recruitment. Improved possibilities for teacher recruitment for the schools should make it possible for schools to improve teacher quality. If students are on average exposed to higher teacher quality, student achievement could be positively affected by the benefits program. I estimate the effect of the program on student achievement in primary and lower secondary education.

The purpose of the reform was to attract more highly qualified teachers. I study if the benefits program was able to do so by estimating the probability to hire more teachers with master's degrees, and more specifically, master's degrees in science or master degrees in pedagogical subjects, teaching degrees or teachers with better academic skills. For student achievement, I study the effects on national test scores in mathematics and reading in 5th and 8th grade, as well as 10th grade exam scores.

To make causal inference, I use a difference-in-differences approach, comparing the difference before and after the reform in schools affected by the reform to the comparison schools in the same time period. The analysis is based on Norwegian register data. The benefits are only granted to the new teachers employed at the school, hence the recruitment effect of improved teacher wages is isolated.

Higher wages could lead to better recruitment by two channels: It could persuade highly qualified individuals to enter the teacher labor market, thus expanding the pool of applicants, or redirect applicants to schools with wage benefits from the other schools. If the effect is a pure sorting effect, there will be compositional changes in teacher recruitment in the surrounding schools. I test whether that is the case, or if the reform expands the local teacher labour force.

For the reform to improve the qualifications of the teacher stock, it is also necessary to retain the teachers. I study whether the turnover pattern of the teachers hired by the reform differ from that of the teachers already hired, and if the teachers already hired react in terms of mobility to the larger wage disparity caused by the reform.

The benefits program consisted of three elements for newly hired teachers in the reform schools: Higher wages, reduced workload in combination with guidance from experienced teachers, and a paied teachers' training course for

teachers lacking formal qualifications to teach, thereby acquiring such qualifications. More than 30 of the schools in the largest school district in Norway have been comprised by the program for at least a year. The implementation of the program started gradually in 2009, and it is still active. The gradual implementation of the reform strengthens the argument for a causal interpretation of the results.

How teachers sort out of schools associated with disadvantaged students and high workload are well documented (see e.g., Jackson, 2009, Bonesrønning et al., 2005 or Lankford et al., 2002). Regarding the effect of wages on teacher composition, the litterature suggest a moderate, positive effect: Ballou and Podgursky (1995) find for exampel a small, positive effect on mean SAT scores on the hired teachers by a 20% raise. Figlio (1997) documents a positive effect of wages on teachers' college selectivity when exploiting variation in relative wages and Gilpin (2012) finds a wage elasticity on aptitude as measured by ACT and SAT scores of .132. Hendricks (2014) studies the effect of wages on teacher retention, and finds that higher teacher pay lead to higher retention rates, and thus longer seniority of the teachers in Texas.

There are some cases of experimental and quasi-experimental methods used to identify the effect of wages on characteristics of the employees. Bó et. al. (2013) exploits a wage experiment to find effects of wages on civil servants in Mexico, and find positive effects on IQ, personality and motivation of the applicants. Falch (2013) use a quasi-natural experiment to estimate the the effect of wages on recruitment of licenced teachers, and find positive effect on recruitment. When using the same experiment to study retention, Falch (2011) finds significant effects of a wage increase on voluntary quits among teachers.

It is not apparent who are the most effective teachers. Rivkin, Hanushek & Kain (2005) discusses how mostly unobservable characteristics contribute to student learning. However, some observables are associated with effectiveness: Experience is suggested to have a positive effect for the first few years after starting a teaching career (see e.g., Rivkin, Hanushek & Kain, 2005 or Harris & Sass, 2011). Hanushek, Piopiunik and Wiederhold (2014) find indications of between-country variation in cognitive ability for teachers (as measured in PIAAC) explaining a substantial part of the variation in PIAAC results for students. Clotfelter et. al. (2010) find that college selectivity and scores on teacher licensure tests have a positive effect on high school performance. Goldhaber & Brewer (1997) and Monk (1994) find that strong background in mathematics or science increase student performance in these subjects. However, teachers with master's degree are not found to be more effective than other teachers (Hanushek, 2006).

If schools can identify their need, attracting a large pool of applicants for available positions increases the probability of finding the right match, which implies possibilities for improved student achievement. Hanushek (2006) argues that even though the applicant pool increases by higher wages, schools are not able to identify the most effective teachers. Naper (2010), on the other hand, finds that schools that do their own hiring hires more effective teachers, indicating that unobservable characteristics of high quality teachers are in fact

observable to the school.

Performance pay has been a focal point of the litterature on teacher wages and student achievement. It is acknowledged to improve student performance, at least in the short run, both by experimental studies (see e.g., Figlio & Kenny, 2007, Lavy, 2009 and Glewwe et. al, 2010) and a cross-country comparison (Woessmann, 2011). On the other hand, when studying teacher selection, performance pay has been suggested to have a negative impact (Falk & Dohmen, 2010).

For the effect of general wage increases on student achievement, there is less evidence. One exception is Loeb and Page (2000), who find that higher wages could lead to decreased high school drop out rates in the US when considering alternative labour market possibilities for the teachers. Dolton and Marcenaro-Gutierrez (2011) exploits cross country variation in relative teacher wages, and find that a wage increase of 10% implies a 5-10% increase in student performance. Incentive-based policies, such as performance pay, have been recommended by e.g., Hanushek and Rivkin (2007) as a more cost efficient way to improve student performance than resource based policies. However, for urban schools who struggle to attract teachers, resource based policies could be considered necessary to counteract teachers sorting out of the vulnerable schools. Machin et. al. (2010) find that more resources in urban schools in London increased student performance in mathematics and attendance.

In this paper, I find positive effects on the probability to hire teachers with advanced degrees, and more teachers with science background. Indication of more teachers with advanced degree in pedagogy are also found. The changes in teacher characteristics do not seem to be a pure sorting effect, but rather due to changes on the extensive margin from the school district's point of view. The results are robust to comparison group.

The results also suggest an effect on student performance. Performance in lower secondary education increases after the reform, as does reading performance in primary education. The reform increases performance in 8th grade reading and on the 10th grade exam by .07 standard deviations of a grade. These findings could suggest that increased wages lead to recruitment of more efficient teachers.

The rest of the paper is organized as follows: Section 2 describes the institutional setting and relevant features of the teacher labor market in Norway. Section 3 outlines the empirical strategy. Section 4 describes the data used and some descriptive statistics. In section 5, the results for teacher recruitment are presented and robustness tests performed, and section 6 presents the results for student achievement. Section 7 offers some concluding remarks.

2 Institutional setting

2.1 The benefit program

The benefit program was introduced in an effort to increase the number of applicants to schools struggling to attract new teachers, with the aim to hire better qualified teachers. The program consists of 3 elements:

- Increased wage of about 5%
- 1 hour of counselling in combination with 1 hour decrease in workload
- Paied teachers' training course for applicants lacking a formal teacher education

7 schools implemented the program in 2009. Most teacher turnover takes place over the summer, and teachers positions starting from august 2009 were announced with the benefits. However, when the new positions were announced for fall 2009, the details of the benefit program were not ready. The job postings only stated that better working terms were to be formulated, and that these would apply to the new employees. It was not yet clear that the better working terms included higher wages. After the first year, the test schools were satisfied with the program, and claimed it increased their number of applicants drastically. The following year, the content of the program was ready before vacancies were advertised. The program expanded to including 15 more schools in 2010, in addition to all the schools from 2009. The schools consist of both primary (age 6-12), lower secondary schools (13-16), and combined primary and lower secondary schools (6-16).

Table 1: Reform schools

	6-12 years	13-16 years	6-16 years	Total
2009	3	2	2	7
2010	12	5	5	22
2011	13	5	5	23
2012	15	5	5	25
2013	16	5	5	26

Table 1 shows that the reform mostly affects primary schools, but also lower secondary schools and combined primary and lower secondary schools are included in the reform schools. The number of schools in the benefits program increases every year.

After 2010, schools move in and out of treatment. The seven schools that joined the program in 2009 are treated for all years. In total, 35 different schools are treated at some point in time, 29 of them for more than one year. The school district authority decides every year which schools should be comprised by the benefits program from the percieved difficulty of recruiting enough qualified teachers.

Schools who are not able to attract the necessary qualified teachers hire personell with low qualifications on short term contracts. The reform thus has the scope to improve teacher quality. Teaching positions will be filled in either case, but the persons filling them will have lower qualifications if the schools are not able to attract enough qualified applicants.

Only the teachers who get hired in a reform school in a reform year receive the benefits, and keep their wage benefits as long as they are hired in that specific school. Reform schools get budget increases to pay for the higher wage expenditures. Both temporary and permanent positions were included in the program.

Whereas almost all teachers hired by the program (more than 90 %) receive the higher wages, not all receive the other benefits. About 2/3 receive counselling, and less than 10 % have started teachers' course. For experienced teachers, counselling is less relevant than for those who lack any teaching experience. Even if the program is aimed at teachers in the start of their career, also more experienced teachers receive the benefits. Moreover, counselling is not necessarily systematic, and could therefore be underreported. As for the low number starting teachers' course, most teachers are not eligible as they alreay have the required pedagogical education.

The program was considered successfull with regards to recruiting teachers, but retaining them has been a challenge. Teacher turnover are high during the first years after entering the teacher labour market. 40% of the teachers employed by the program have left the school two years after starting. There are a few cases of reported discontent of the remaining teachers at the schools, but 73% of the schools report no discontent.

2.2 Teachers

Teacher wages are strictly regulated in Norway, and exhibits very little between-school variation. Teacher wages are almost entirely decided from educational backgroun and seniority. Other important factors are also fixed; There are no variation in work hours or fringe benefits. Hence, small changes in wage could change the sorting patterns.

Most Norwegian teachers have a dedicated teacher education, consisting of a four year program. The alternative is higher education degrees supplemented with a teachers' course, especially common for lower secondary education teachers. Both educational backgrounds ensures eligibility for permanent positions. Also persons without teacher education work as teachers at schools, but are only allowed to be hired on short term contracts.

We are not able to distinguish teachers in permanent positions from others. Most teachers in the reform schools start with short term contracts, and are later appointed to permanent contracts. Teachers with both permanent and temporary contracts are included by the program.

2.3 Student performance

In Norway, there are no grading in primary education. In 2007, national testing was introduced at age 10 and 13. From 2010, national testing are performed at age 14 as well. National testing are low stakes for the students, as they face no consequences for low performance.

In lower secondary education, teacher grading is the main grading practice. Teacher grades are given in a total number of 13 subjects, and there is one central exit exam. The grading on the exam is external, and the students sit for the exam in either English, mathematics or Norwegian. The external grading makes the exam grade credible, but also serves as a benchmark for the teacher grades to be credible. The mean of all grades, including exam grades, are used to compete for seats in upper secondary education, and are thus high stakes for the students.

3 Empirical strategy

3.1 Teacher recruitment

To estimate a causal effect of the reform on teacher recruitment, there are two elements that need to be taken into account. First, there may be variation over time in teacher recruitment that are unrelated to the benefits program. Random variation in the exit rates from the schools will increase the demand for teachers in a school while size and composition of the graduating cohort from higher education may affect the supply of potential teachers. Second, reform schools are included in the program as a response to difficulties with recruiting qualified teachers, and are thus by definition different from schools not included. Different demand for teachers are therefore not surprising.

For a causal interpretation, it is necessary to identify the effect on the characteristic in question of being hired by a reform school in a reform year. I compare the difference between reform schools and non-reform schools before and after the reform.

Whereas turnover has been widely studied, recruitment is a more challenging outcome. The number of applicants to vacant positions are not observable, only the actual hirings. The outcome studied in this paper is the probability to hire a teacher with certain characteristics.

The following equation is estimated using a difference-in-differences methodology:

$$y_{st} = \alpha + \beta x_{st} + \gamma treat_s + \mu year_t + \theta(d_{st}) + \varepsilon_{st} \tag{1}$$

Equation 1 is our easiest specification, where y_{it} is the probability of hiring a person with the characteristic in question at time t for school s. The effect on the following characteristics are estimated:

- Advanced degree (more than four years of higher education)
  - Advanced degree in science

  - Advanced degree in pedagogical subjects

- Teaching degree
- Teachers' academic skills

All the studied characteristics are measures of qualifications. The object of the reform was to recruit more qualified teachers, and educational traits are easily observable to the employer.

Included in control variables x_{st} are controls for grade level at the school: Age 6-12, 13-16 or 6-16, or school fixed effects. $treat_s$ is a dummy for schools being treated at some point in time, $year_t$ is a vector of year fixed effects, and d_{st} is a dummy for treatment status for school s in year t. Our variable of interest is thus θ, which expresses the change in teacher characteristics at a reform school in a reform year. The error term ε_{st} is clustered at school level to account for correlated error terms within schools. In the main specification, schools in the rest of the country is defined as the comparison group.

To study the mechanisms behind the estimated effects further, I estimate the effect using the surrounding schools as a control group, and when excluding the surrounding schools. If the mechanism behind the reform is to redirect applicants from nearby schools, we should observe a proportional change in recruitment in nearby schools. By comparing the estimates when only including nearby schools in the comparison group and when excluding them, I check if the estimated reform effect is caused by changes on the extensive margin, i.e., more people go into the teaching profession, or if what we see is strictly a sorting effect.

The critical assumption in a differences-in-differences approach is that in absence of the reform, the treatment and comparison group would follow a similar trend. Because of relatively few observation in the treatment group, the time trends for the outcomes are volatile, and to graphically assess the common pre-trend assumption is challenging. I therefore perform a placebo test. I estimate year fixed effects for the treatment group both before and after treatment and test for any reform effects in the pre-reform years.

A concern when applying a differences-in-differences approach is that schools could be affected differently by common shocks. In our case, the benefits program was introduced in different years for different schools. The gradual implementation makes the empirical approach more robust to such shocks as the reform status varies over time between the schools.

The reform was a response to difficulties with recruiting teachers, and treatment is thus not randomly assigned, as would be the ideal. I exploit the gradual implementation of the reform, as schools included later in the period face similar recruitment difficulties as the ones included first. . I test whether the results are robust to using schools where implementation took place in 2011 or later as a comparison group when estimating the reform effect in 2009 and 2010.

Each school district are in charge of their own policies. If Oslo has changed general school policies in the reform years, what I interpret as the reform effect may actually be attributed to being located in the school district of Oslo. In

addition, Oslo is a large city, and may be exposed to different labour market shocks than other parts of the country. To safeguard against this, I also use Oslo as a comparison group.

The reform schools had all experienced problems with attracting applicants for their available positions, and are different than other schools in the school district. If there are trends or shocks that affect these schools differently than other schools, the results may be biased. On the other side of the school district border, the schools are similar to the schools on the treatment side of the border. They belong to the same labour market area, and are thus exposed to the same labour market shocks as the reform shools. I use schools in the surrounding school districts as a comparison group as a robustness test.

If labor market conditions are more generally related to proximity to large cities, other larges cities is a relevant comparison group. I repeat the analysis using only large cities[1] as comparison group, which could have easier access to high skilled labour.

The indicator of teachers' academic skills are based on the grades obtained in the highest degree in higher education. I check if the results are robust to different specifications of the indicator: All grades in higher education, only grades in pedagogical subjects, only grades in mathematics and science, or using a dummy for having education from a selective institution.

3.1.1 Further analysis

For the schools to increase the qualifications of their teacher stock, it is necessary also to retain the teachers. To study if teachers hired by the reform differ in terms of turnover, the following equation is estimated:

$$y_{ist} = \alpha + \beta x_{it} + \eta sfe_s + \gamma treat_i + \mu year_t + \theta treatment_i + \varepsilon_{ist} \qquad (2)$$

I estimate the probability for person i in school s at time t quit the teaching job. We control for school fixed effects sfe_s and year fixed effects $year_t$. Included in x_{it} are educational background, gender, seniority and age. Our variable of interest θ measures the effect of being hired in a reform school in a reform year on turnover.

To see whether the introduction of the reform had any implications for teacher turnover for teachers not comprised by the reform, but who are working at the reform schools in a reform year, I estimate their propensity to quit. I exclude those who received the benefits, and define treatment as working in a reform school in a reform year, but not receiving the benefit. Equation 2 is then estimated with the alternative treatment definition, but the variable of interest is still θ.

[1] Bergen and Trondheim, in addition to rest of Oslo

3.2 Student performance

To study how the benefits program affect student performance, I use a similar empirical strategy as for teacher recruitment.

The following equation is estimated:

$$y_{ist} = \alpha + \beta x_{is} + \gamma treat_i + \mu year_t + \theta(d_{it}) + \varepsilon_{ist} \tag{3}$$

Our outcome y_{it} is student performance for individual i at time t in school s, measured by national test scores in mathematics and reading for 5th and 8th grade, and exam results from 10th grade.

x_{is} are control variables for gender and school, $treat_s$ is a dummy variable taking the value 1 for schools in the treatment group, and $year_t$ indicates year fixed effects. d_{st} is a reform indicator that equals 1 if the school has been included in the reform in the same or an earlier year. Our parameter of interest is again θ, expressing the estimated effect of being in a reform school in a post-reform year. Standard errors ε_{ist} are clustered at school level.

Previous knowledge are key to student performance. I include controls for test scores at the previous level for the cohort who takes the test in question at time t at school s. By controlling for earlier student achievement at the school level, the empirical approach is less sensitive to measurement error than if we were to control for earlier student achievement at the individual level. Test scores at the previous level is available for 8th grade and 10th grade.

Learning is accumulative, and does not only depend on changes occuring the former year. In Equation 3 reform effects are pooled together in θ independant of how long the students have been exposed to the reform.

To check whether the common trend assumption holds, I perform placebo tests also for student achievement. I study if there are significant differences in pre-treatment trend in the treated schools and comparison schools by estimating the effect of being in a reform school for each year.

When studying performance in 8th grade, there are no pre-treatment period. I am thus not able to perform placebo tests for 8th grade performance.

A concern for the student achievement analysis is that student composition are different in the treatment and the comparison group, and that external influences may have different impact according to social and immigrant background. For student achievement, later implementation schools are a less optimal comparison group than for teacher recruitment. Learning is a cumulative process, and I check if the results are due to Oslo-specific trends by only including other schools in Oslo as the comparison group.

There could be differences in how the school approaches national testing. I test this by only including students who sit for both tests in the same school in the sample, i.e., only student in 1.-10. grade schools.

4 Data

4.1 Teacher recruitment

When estimating the effect of higher wages on teacher recruitment, we use administrative register data from Statistics Norway. Our main source to define employment spells is the employer-employee register (Arbeidstakerregisteret). Every employment spell is registered, so we are able to follow teachers throughout their career. There are information about start of employment, positions, school identifier and manhours worked. There are also information about personal characteristics, such as age, gender and educational background. I use data from 2005 to 2012.

I use data on academic achievement for the teachers (FS-registeret). All grades from higher education are registered, and are used to estimate an indicator of academic achievement. Grades are normalized within year, field of study and institution, and the indicator are calculated from all registered, normalized grades. The main indicator is the mean of all normalized grades in higher education.

When studying recruitment, only the teachers starting in a new job are of interest. The sample consists of all individuals who start in a teaching position at a new school between age 20 and age 60. Only teachers working at least half time, i.e., more than 20 hours per week are included. Starting in a teacher position is defined as being registered as employed in a school where the person did not work the year before. This leaves us with 89 957 teachers starting in a new school distributed over 3 282 schools, of which 2 488 teachers in the reform schools.

Figure 1: Teacher mobility

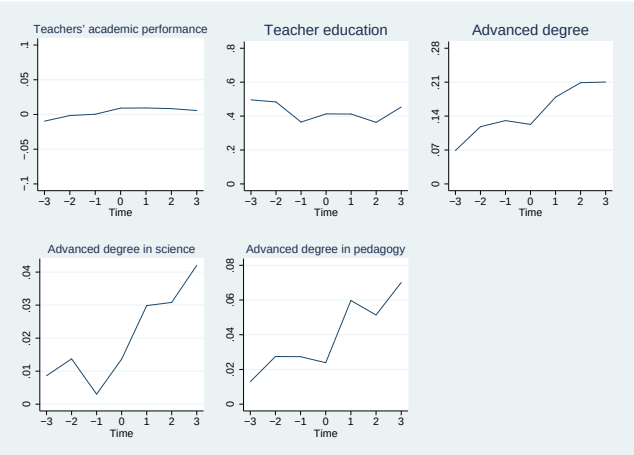
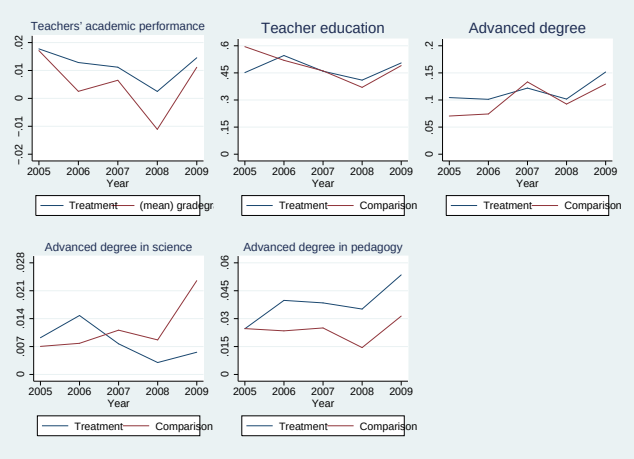


Figure 1 shows the proportion of hirings with the characteristics in question both for the treatment group. There are no indications of a post-reform change

Figure 2: Pre-reform trends



for academic achievement and share with formal teacher education. For teachers with advanced degrees, figure 1 shows an increase in the hiring probability after the reform. Both the share with advanced degrees in science and pedagogy increases after the reform.

Although Figure 1 shows increased hirings of teacher with advanced degrees after the reform, it does not show the hirings for schools not affected by the reform. Figure 2 shows the hirings for reform schools and the other schools in Oslo from 2006 to 2009.

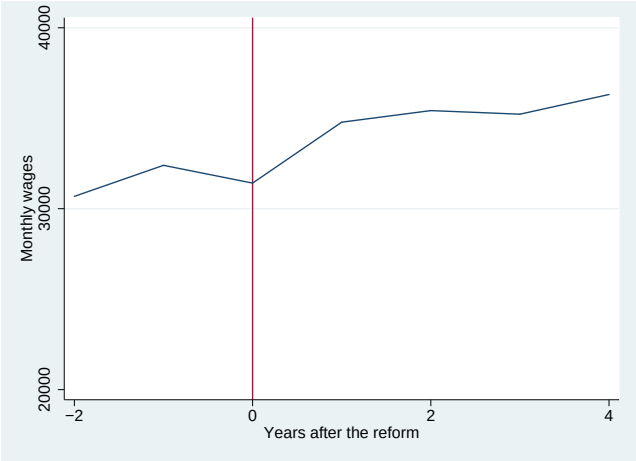
The academic achievement and the share with formal teacher education are similar for reform schools and the other schools in Oslo. For the share with master degrees in science, there is more variation. There is an increase in 2009 for the reform schools, which is the first year of the reform for seven of the schools schools. The numbers of teachers with master degrees in science are small, so small variations in the number of persons with the characteristic have large impact on the share.

The main content of the reform was increased wages, and a discontinuity in the wages of the newly hired should be possible to spot in the data. Table 3 indeed shows a discontinuity in the monthly wages of newly hired teachers in the reform schools at the time of implementation. [2]

Table 2 shows how the reform schools differ from non reform schools. Most interestingly is that the share with teaching degree are lower in the treatment schools, and the share with low education, i.e., upper secondary education or lower, are higher in the reform schools than in the comparison group. All measures of student performance is substantially lower in the treatment schools. For more general school characteristics, the schools are larger with more students,

[2]Included in Figure 3 are all new teachers who start in a teaching position in a reform school the years before and after the reform. Only teachers starting in august are included for comparability across years.

Figure 3: Wages of the Newly Hired Teachers in the Reform Schools



and the resources spent on special education, including language instruction, are substantially higher. The high special education index indicates higher share of minority students.

Table 2: School characteristics

	2009 implementation	2010 implementation	Implementation after 2011	Comparison group
Teacher characteristics				
Advanced degree	13.08	6.85	8.18	5.83
Teaching degree	43.46	56.01	43.03	56.93
Low education	24.62	19.48	27.27	16.55
Academic achievement	.008	.012	-.004	.009
Student Performance				
5th grade mathematics	-.534	-.230	-.122	.005
5th grade reading	-.625	-.253	-.177	.006
10th grade exam	-.287	-.238	-.067	.042
Student composition				
Number of students	385	417	336	219
Special education index	.687	.710	.834	.112

4.2 Student performance

Student performance in primary education is measured by National test scores from 5th and 8th grade. I have test score data from 2007 to 2014. Test scores are standardized within subject and year, with mean zero and standard deviation of

one. Any year to year variation and systematic differences in grading practices by subject are thus accounted for.

Central exit exams are used to measure performance in lower secondary education. I normalize the grades within subject and year to account for any differences in grading practices between subjects and year to year variation.

Figure 4: Student performance pre-reform

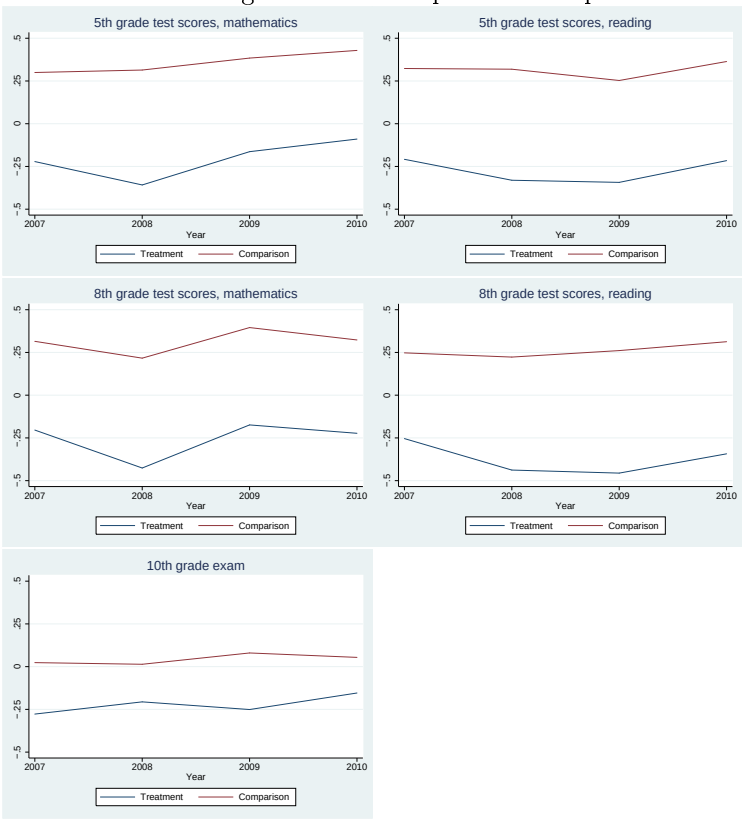


Figure 4 shows student achievement from 2007 to 2010 for 5th, 8th and 10th grade in the reform schools, and in the rest of the schools in the same school district. The schools follow a similar pattern pre-reform. The existence of common trend suggest that there are factors within the school district that affects students similarly. For 10th grade, there seem to be a slight increase in the treatment area in 2010. Treatment are implemented in the first schools in 2009, so there is a possibility of treatment effects in 2010.

5 Results for teacher recruitment

5.1 Main results

Table 3 shows the estimated effects on teacher characteristics of the reform by equation 1 with school fixed effects included.

Table 3: Teacher recruitment

	Treatment effect	Treatment effect with school fixed effects	Mean, treatment group	Mean, comparison group
Advanced degree	.0721*** (.0200)	.0691 (.0179)***	0.1362	0.0735
Advanced degree in science	.0249*** (.0082)	.0200 (.0080)**	0.0182	0.0082
Advanced degree in pedagogy	.0248** (.0122)	.0269 (.0120)**	0.0364	0.0224
Teaching degree	-.0232 (.0488)	-.0123 (.0421)	0.4475	0.5472
Academic achievement	.0016 (.0047)	.0023 (.0050)	0.0056	0.0088

Note: All specifications include a constant term, year dummies (ref. 2005), and the a dummy for treatment status $treat_s$. The second specification is used in all subsequent tables on teacher mobility, all estimated by OLS. Standard errors are clustered on school level. */**/*** statistically significant at the 10/5/1 percent level.

Table 3 shows a strong effect on the probability to hire teachers with master degrees after the reform, and more specifically, an increase in the probability to hire teachers with master's degree in science. I find no effect on the probability to hire teachers with a designated teaching degree, and no effect on the hired teachers' academic achievement. When adding school fixed effects, the estimates do not change.

To study if the compositional changes in the hired teachers are due to changes on the extensive or intensive margin, I compare the estimates when using schools in the proximity and schools further away. If the improved quality is due to internal sorting of the teachers, a stronger effect will be found when using the nearby schools than schools further away. The reform will not only lead to increased probability of hiring persons with certain characteristics, but also decreased probability of hiring the individuals with the same characteristics in the surrounding schools.

Table 4: Sorting or expanded pool of applicants

	Same labor market	Different labor market
Advanced degree	.0390** (.0171)	.0573*** (.0175)
Advanced degree in science	.0170** (.0077)	.0207*** (.0078)
Advanced degree in ped.	.0120 (.0104)	0.0212** (.0101)
Teaching degree	.0045 (.0408)	-.0088 (.0419)
Academic achievement	.0064 (.0053)	.0032 (.0050)
N	22 563	71 578

Note: See Table 3

In our comparison of the intensive or extensive margin in Table 4, the higher point estimates when using schools outside the same labor market area for share with advanced degree are higher than when only using the surrounding area as a comparison group. The same is the case for both the subgroups of advanced degrees we study. For the share with advanced degrees in pedagogy, the estimated effect becomes non significant when using the same labour market as comparison group. For the share with teaching degree and for academic achievement, the point estimates are somewhat higher in the surrounding area, but not significantly so. I therefore conclude that there are no indications of the estimated effect being purely a sorting effect.

5.2 Robustness tests

5.2.1 Placebo test

Table 5: Placebo test teacher recruitment

Years	Master degrees	Science degrees	Pedagogical degrees
-7	.0667 (.0673)	-.0072 (.0068)	.0075 (.0093)
-6	-.0464 (.0551)	-.0066 (.0064)	.0006 (.0098)
-5	-.0023 (.0339)	-.0025 (.0060)	.0195 (.0191)
-4	-.0082 (.0273)	-.0044 (.0085)	.0232* (.0120)
-3	-.0355 (.0215)	-.0034 (.0087)	-.0041 (.0095)
-2	.0188 (.0262)	.0024 (.0090)	.0118 (.0127)
-1	.0188 (.0295)	-.0087 (.0078)	.0059 (.0121)
1	.0559* (.0294)	.0168 (.0140)	.0329** (.0155)
2	.0714** (.0346)	.0154 (.0128)	.0198 (.0188)
3	.0655** (.0296)	.0233 (.0160)	.0376* (.0226)
4	-.0025 (.0318)	.0244 (.0152)	-.0102 (.0121)
5	-.0214 (.0516)	-.0377*** (.0124)	.0687** (.0339)
R2	.0835	.0472	.0486
N	96 590	96 590	96 590

Note: See Table 3

Table 5 shows that the effect of being in a reform school before and after implementation of the reform. For the share with advanced degrees, there are no significant reform effects before the reform. The same is the case for science degrees. The point estimate four years before the implementation is significantly positive for the share with master degree in pedagogy. However, the characteristics of the new recruited teachers are volatile. The placebo test reveals a pattern with fluctuations around zero before implemetation of thereform, and higher estimates after the implementation.

5.2.2 Comparison groups

I exploit the gradual implementation of the reform by using schools where treatment not yet has been introduced as comparison schools when identifying the 2009-2010-effect. The benchmark is the reform effect when using all non-treated schools are comparison group in the estimation of effects in 2009 and 2010. The effects are smaller when estimating the reform effect only in 2009-2010 for the probability to hire teachers with advanced degrees and science degrees., and for advanced degrees, significantly so. The estimate for advanced degree in pedagogical subjects is no longer significant when only including later implementing schools in the comparison group.

Table 6: Reform schools as comparison group

	Benchmark	Later implementation schools as comparison	Difference
Advanced degree	.098*** (.022)	.064** (.029)	.047*** (.017)
Advanced degree in science	.022** (.011)	.021* (.011)	.011 (.005)
Advanced degree in ped.	.035** (.014)	.024 (.023)	.008 (.010)
Teaching degree	-.067 (.048)	-.017 (.045)	-.040 (.047)
Academic achievement	.008 (.006)	.001 (.011)	.001 (.007)

Note: See Table 3

Table 7 shows that the results for teacher recruitment are not sensitive to choice of comparison group for most outcomes. Large cities are thought of as having easier access to high skilled labour than less densely populated areas, but also as having more options on the labour market. Using large cities as comparison group do however not alter our results, unless for the share with advanced degrees in pedagogy. In row 3, we see the treatment effects when using schools in the same labour market area. The comparison school district do not have a similar benefit program as in the treatment district, and the student composition are similar on each side of the school district border. The results are slightly decreased.

When only using schools within the same school district as comparison group, we make certain that the estimated effect is not driven by other school district specific policy. The estimates is reduced, but remain significant at 10% level, except for the share with advanced degree in pedagogy. The effect is thus not an effect of being in in that specific school district.

5.2.3 Indicator of academic achievement

In the main analysis, the indicator of academic achievement that is based on grades from the highest completed degree in tertiary education. I test if the results are sensitive to the choice of indicator, and estimate the effect on a series of indicators: Achievement only in pedagogical subjects, achievement in science and mathematics, all grades in higher education, and selective schools information. I find no effects on any indicator when estimating equation 1.

Table 7: Comparison groups

	Benchmark	Large cities	Surr. school district	Oslo
Advanced degree	0.0535 (0.0173)***	0.0421 (0.0172)**	0.0390 (0.0171)**	0.0323 (0.0176)*
Advanced degree in science	0.0197 (0.0077)**	0.0178 (0.0078)***	0.0171 (0.0077)**	0.0140 (0.0079)*
Advanced degree in ped.	.0193* (.0101)	.0108 (.0105)	.0120 (.0104)	.0074 (.0109)
Teaching degree	-0.0062 (0.0418)	-0.0295 (0.0416)	0.0045 (0.0408)	-0.0010 (0.0404)
Academic ach.	0.0040 (0.0049)	0.0018 (0.0052)	0.0064 (0.0053)	0.0025 (0.0055)

Note: See Table 3

Table 8: Quit behaviour

	Quit behaviour for teachers receiving the benefits	Quit behaviour for teachers in reform schools who do not receive the benefits
Treatment	-.0555*** (.0186)	-.0222 (.0144)
R2	.3509	0.1360
N	585 173	582 895

Note: See Table 3

5.3 Quits

I test if teachers hired by reform schools in the reform period has different turnover behaviour than teachers not affected by the reform.

Table 8 shows that reform teachers have a lower turnover rate compared to teachers hired by the reform schools before the reform was implemented. As controls for age and seniority at the school are included, it is not due to compositional changes of the teacher stock in the reform schools. The turnover behaviour of the reform teachers indicates that they are different from the newly hired teachers hired before the implementation of the reform.

For teachers not comprised by the reform, but already hired in the reform schools, there are no find evidence of increased turnover in the reform years. Although some discontent for teachers in reform schools not receiving the benefits was reported, it did not result in any increased turnover. It can also be ruled out that teachers hired in the reform schools quit their jobs to apply for teaching in jobs in other reform schools, and thereby gaining the wage increase.

6 Results for student performance

6.1 Main results

To study how the benefits program impact student performance, I estimate the effect of the reform as in Equation 3. The outcomes for student performance are test scores from 5th and 8th grade national testing in mathematics and reading, and exam scores from 10th grade.

Table 9: Student performance

	DD	DD, with school fixed effects	DD, with school fixed effects and controls for earlier achievement	Mean, treatment district
Mathematics 5th grade	.0069 (.0648)	.0335 (.0601)		-.2203
Reading 5th grade	-.0201 (.0485)	.0193 (.0505)		-.2897
Mathematics 8th grade	.0399 (.0484)	.0186 (.0282)	.0996*** (.0358)	-.2491
Reading 8th grade	.0722 (.0506)	.00983** (.0380)	.1898*** (.0430)	-.3385
10th grade Exam	.0509 (.0327)	.0632** (.0285)	.1434*** (.0237)	-.2353

Note: See Table 3

Table 9 shows no effects of the reform on student performance when not including school fixed effects. When adding school fixed effects, point estimates are slightly increased and standard errors decreased, and a positive effect appears for 10th grade exam results and for 8th grade test scores in reading. The last specification includes controls for the mean achievement at the cohort level at the former test, and shows substantial positive effects for performance in both 8th and 10th grade, doubling the point estimates from column 2. In the following robustness tests, I will use the specification from column 2.

The effects are substantial given that he reform did only apply to the newly hired teachers and did not give any incentives for increased effort. However, the schools could strategically assign the new, well qualified teachers to the students with the most to gain, who in lower secondary education is likely to be students in their last year. Moreover, Jackson and Bruegman (2009) have found evidence of peer effects among the teaching staff in schools: More effective collegues induces a higher value added for a given teacher.

6.2 Robustness tests

6.2.1 Placebo tests

Table 10 shows the estimated effects of the reform before implementation. For 5th grade reading, there is a positive placebo effect at 10% level three years before implementation, showing the volatility of the results in the treatment group. There is a similar finding for 8th grade mathematics two years before implementation. There are no significant placebo effects for any of the subjects where there are reform effects.

The placebo test reveals an interesting pattern. There is a a cumulative effect with increasing student achievement after implementation. Students who are measured after a long time are both exposed to the potential high quality teachers for a longer time, and there are more of the new teachers hired at the school, thus increasing the probability of being exposed to them.

Table 10: Placebo test student achievement

	5th grade		8th grade		10th grade Exam
Years	Mathematics	Reading	Mathematics	Reading	
-3	.022 (.108)	.154* (.090)	-.045 (.077)	.0002 (.097)	-.009 (.085)
-2	-.132 (.096)	.029 (.073)	-.178* (.103)	-.121 (.105)	.021 (.066)
-1	-.013 (.090)	-.025 (.081)	-.073 (.090)	-.191 (.084)	.031 (.068)
0					
1	-.001 (.085)	-.015 (.070)	.044 (.060)	.001 (.046)	.018 (.062)
2	.0045 (.093)	-.031 (.084)	-.129 (.060)	-.125 (.048)	.084 (.076)
3	.014 (.100)	.143 (.092)	-.175 (.091)	-.099 (.065)	.060 (.065)
4	.056 (.095)	.218*** (.055)	.024 (.069)	.051 (.066)	.129* (.070)
5	-.075 (.130)	.178* (.101)	.282*** (.100)	.182* (.097)	.132 (.104)
R2	.081	.066	.077	.079	.073
N	455 437	449 772	320 957	318 253	451 845

Note: See Table 3

Table 11: Later implementation schools as comparison group

	Benchmark	Later implementation schools as comparison group	Difference
Mathematics 5th grade	.480*** (.150)	.387*** (.182)	.118 (.085)
Reading 5th grade	.309*** (.088)	.288** (.120)	.072 (.069)
Mathematics 8th grade	-.122 (.152)	-.216 (.271)	.050 (.057)
Reading 8th grade	-.097 (.185)	-.211* (.235)	.036 (.042)
10th grade exam	.022 (.028)	-.099 (.067)	.112** (.054)

Note: See Table 3

6.2.2 Comparison groups

I check whether the results for student achievement are robust to different comparison groups.

The benchmark case are estimated with the rest of the country as comparison group on data up to 2010. Table 11 compares it to the results on student achievement when using schools that implemented the reform in 2010 as comparison group. The difference between the estimates are not significant, with the exception of 10th grade exam results. For 10th grade achievement, the later implementing schools are few, only 6 schools are included in the comparison group. The results are therefore volatile.

There is a clear challenge with this test. It is only possible to estimate the immidiate effect, i.e., the effect on tests taken one year after the first teachers started. Results from the placebo tests in Table 10 reveals a tendency that the effect on student achievement increase with exposure time.

Table 12 shows that the estimated results are robust to changing the com-

Table 12: Comparison groups, student achievement

	Benchmark	Big cities	Surrounding school district	Oslo
Mathematics 5th grade	.034 (.060)	.042 (.062)	.036 (.061)	.038 (.062)
Reading 5th grade	.019 (.051)	.002 (.053)	.009 (.052)	-.002 (.054)
Mathematics 8th grade	.019 (.028)	.060* (.034)	.062* (.036)	.067* (.036)
Reading 8th grade	.098*** (.038)	.095** (.045)	.116*** (.044)	.098** (.048)
10th grade exam	.063** (.029)	.076** (.034)	.053 (.034)	.066* (.036)

Note: See Table 3

parison group. For 8th grade reading and exam performance, both big cities as comparison group and surrounding school district increases the point estimates somewhat, and thus the significance level. When using large cities and surrounding school district as comparison group for 8th grade reading, there are significant reform effects also when not including former cohort performance.

7 Concluding remarks

This paper studies the effect on teacher recruitment of a wage increase only affecting newly hired teachers in chosen schools in Norway. I find evidence of a wage increase of about 5% increases the probability of hiring teachers with master degrees, and teachers with master degrees in science. The results also suggest a positive effect for teachers with advanced degrees in pedagogy. No reform effects are found for the probability to hire teachers with a designated teaching degree or improved academic achievement.

There are no indications of the improved teacher force to be at the expense of the surrounding schools, but rather due to changes on the extensive margin from the school district's point of view. I find no evidence of decreased probability to hire teachers with the characteristics in question in the surrounding schools.

Although high turnover is considered a challenge for the schools, the teachers hired by the reform have a significantly lower turnover than comparable teachers. The remaining teachers in the reform schools who did not receive the benefits do not react to the increased within-school wage disparity in terms of increased turnover.

When studying the effect of the wage increase on student achievement, I find evidence of increased performance in 8th grade reading and performance in lower secondary education. No effect is found for mathematics, which is at odds with the positive findings for the probability to hire teachers with master degrees in science. A possible explanation could be that mathematics is of a more accumulative nature than e.g., reading, and thus requires a longer time to improve the results.

References

[1] Dale Ballou & Michael Podgursky. (1995). "Recruiting Smarter Teachers." The Journal of Human Resources. 30 (2) 326-338

[2] Ernesto Dal Bó, Frederico Finan & Martín A. Rossi. (2013). "Strengthening State Capabilities: The Role of Financial Incentives in the Call to Public Service." The Quarterly Journal of Economics. (2013) 1169-1218

[3] Hans Bonesrønning, Torberg Falch & Bjarne Strøm. (2005). "Teacher Sorting, Teacher Quality, and Student Composition." European Economic Review, 49 (2005) 457-483

[4] Charles Clotfelter, Helen Ladd, and Jacob Vigdor. 2006. "Teacher-Student Matching and the Assessment of Teacher Effectiveness." Journal of Human Resources 41(4):778–820.

[5] Charles Clotfelter, Helen Ladd, and Jacob Vigdor. 2007. "Teacher Credentials and Student Achievement: Longitudinal Analysis with Student Fixed Effects." Economics of Education Review 26(6):673–82. December.

[6] Charles Clotfelter, Elizabeth Glennie, Helen Ladd & Jacob Vigdor. (2008). "Would higher salaries keep teachers in high-poverty schools? Evidence from a policy intervention in North Carolina." Journal of Public Economics, 92(5), 1352-1370.

[7] Charles Clotfelter, Helen Ladd, and Jacob Vigdor. 2010. "Teacher Credentials and Student Achievement in High School. A Cross-subject Analysis with Student Fixed Effects." The Journal of Human Resources, 45 (3) 655-681.

[8] Thomas Dohmen and Armin Falk. 2010. "You Get What You Pay For: Incentives and Selection in the Education System." The Economic Journal, 120 (546) 256–271

[9] Torberg Falch. (2011). "Teacher mobility responses to wage changes: Evidence from a quasi-natural experiment." American Economic Review. 101 460-465.

[10] Torberg Falch (2013). "Wages and Recruitment: Evidence from External Wage Changes." CESIFO Working Paper 4078

[11] David Figlio. (1997). "Teacher salaries and teacher quality." Economics Letters, 55 (2) 267-271.

[12] David N. Figlio, Lawrence W. Kenny. 2007. "Individual teacher incentives and student performance." Journal of Public Economics, 91 (5–6), 901–914

[13] Gregory A. Gilpin. (2012). "Teacher salaries and teacher aptitude: An analysis using quantile regressions." Economics of Education Review, 31 (3) 15-21

[14] Paul Glewwe, Nauman Ilias, and Michael Kremer. (2010). "Teacher Incentives." American Economic Journal: Applied Economics 2 (July 2010): 205–227

[15] Goldhaber, D. & Brewer, D. (1997). "Why Don't Schools and Teachers Seem to Matter? Assessing the Impact of Unobservables on Educational Productivity." Journal of Human Resources, 32(3), 505-523.

[16] Eric A. Hanushek. (2006). "School Resources." Handbook of the Economics of Education, Volume 2, pp. 865-908.

[17] Eric Hanushek & Steven G. Rivkin. (2006). "Teacher Quality." Handbook of the Economics of Education, Volume 2, pp. 1052-1078.

[18] Eric Hanushek, Marc Piopunik & Simon Wiederhold. (2014). "The Value of Smarter Teachers: International Evidence on Teacher Cognitive Skills and Student Performance." NBER Working Paper 20727

[19] Douglas N. Harris & Tim R. Sass. (2011). "Teacher training, teacher quality and student achievement." Journal of Public Economics 95 (2011) 798-812

[20] Matthew D. Hendricks. (2014). "Does it pay to pay teachers more? Evidence from Texas." Journal of Public Economics 109 (2014) 50-63

[21] Kirabo Jackson. (2009). Student Demographics, Teacher Sorting and Teacher Quality: Evidence from the End of School Desegregation." Journal of Labor Economics. 27 (2) 213-256.

[22] Jackson, C. K., and E. Bruegmann. 2009. Teaching students and teaching each other: The importance of peer learning for teachers. American Economic Journal: Applied Economics 1(4): 85-108.

[23] Lankford, H., Loeb, S., & Wyckoff, J. (2002). Teacher sorting and the plight of urban schools: A descriptive analysis. Education Evaluation and Policy Analysis, 24(1), 37-62.

[24] Victor Lavy. (2009). "Performance Pay and Teachers' Effort, Productivity, and Grading Ethics." American Economic Review, 99(5): 1979-2011.

[25] Stephen Machin, Sandra McNally & Costas Meghir. 2010. "Resources and Standards in Urban Schools." 4 (4) 365-393

[26] David H. Monk. (1994). "Subject Area Preparation of Secondary Mathematics and Science Teachers and Student Achievement." Economics of Education Review, 13, 2: 125-145.

[27] Linn R. Naper. (2010). "Teacher hiring practices and educational efficiency." Economics of Education Review, 29, 4: 658-668

[28] Steven G. Rivkin, Eric Hanushek & John F. Kain. (2005). "Teachers, Schools and Academic Achievement." Econometrica 73 (2) 417-458

[29] Tim R. Sass, Jane Hannaway, Zeyu Xu, David N. Figlio & Li Feng. (2012). "Value added of teachers in high-poverty schools and lower poverty schools." 72 (2012) 104-122

[30] Rothstein, Jesse. 2010. "Teacher Quality in Educational Production: Tracking, Decay, and Student Achievement." Quarterly Journal of Economics, 125(1): 175–214.

[31] Ludger Woessmann. 2011. "Cross-country evidence on teacher performance pay." Economics of Education Review 30 (2011) 404–418

Appendix