Gender differences in competitiveness: Evidence from educational admission reforms*

Arnt O. Hopland† and Ole Henning Nyhus‡

April 15, 2013

Abstract

Using survey data on student motivation, this paper studies if increased competition has adverse consequences for students' intrinsic motivation and, in particular, if there are gender differences. In recent years, two Norwegian counties have introduced stronger competition among students in lower secondary education through reforms of their upper secondary education admission systems. Using a difference-in-differences approach we find that whereas the motivation for boys seems to be unaffected by the increased competition, there are clearly adverse consequences on the motivation for girls.

JEL classification: D02, D03, I20, I28, J16

Keywords: Gender differences, competitiveness, student motivation

*Valuable comments from Fredrik Carlsen, Torberg Falch, Marianne Haraldsvik, Bjarne Strøm, and participants at the 2013 EPCS Meeting (Zurich), 2013 Annual Meeting of the Norwegian Economic Association (Stavanger), FIBE 2013 (Bergen), a workshop in Lisbon and seminar participants in Trondheim are gratefully acknowledged. All remaining errors are our own.

†Department of Economics, Norwegian University of Science and Technology. Email: arnt.hopland@svt.ntnu.no

‡Department of Economics, Norwegian University of Science and Technology. Email: ole.nyhus@sof.ntnu.no

1 Introduction

Experimental economists have identified significant gender differences in the willingness to compete as well as how outcomes are affected by competition.[1] The general conclusions are that males are more willing to compete than females, and that males also perform better than females in competitive settings. Such differences may have the potential to explain gender differences in outcomes in a variety of settings, e.g. in the labor market. In recent years, this literature has been supplemented by a few field studies which in general provide more mixed results.[2] Hence, more empirical evidence on gender differences in competitiveness, should be of great interest.

Our empirical analysis is a large-scale field study where we utilize data from upper secondary education admission reforms in two Norwegian counties. The counties decide how students are allocated between the schools, and an important issue is whether admission to a specific upper secondary school is based on the neighborhood principle or performance based ranking (school choice). Whereas the prior involves little competition between students, since students enroll in the neighboring school, the latter introduces competition between lower secondary students with ambitions of entering an attractive upper secondary school.

In the empirical analysis, we make use of survey data on student motivation collected by the Norwegian Directorate for Education and Training and admission reforms in two counties, Oslo and Hordaland. Both counties changed from an admission system based on the neighborhood rule, to a system with performance based ranking. The reforms did, however, take place at different points in time. The reform was implemented in Hordaland in the school year 2004/05 while it took place in Oslo in the school year 2009/10. We will study the two reforms separately, but their similarities allow us to use identical empirical strategies when analyzing them.

Large-scale field data, such as ours, have several desirable properties compared to experimental data. The first is related to external validity. The data is tapped from a survey covering several cohorts in Norwegian lower secondary school, and reforms affecting all students in two Norwegian counties. Hence, whereas experimental studies usually have to rely on small and often not very representative samples, our analysis is indeed based on a large and representative sample. Second, the reforms are of real importance for students' possibility to be enrolled in their upper secondary school of choice. The stakes are thus clearly of a non-negligible character, rather than small or even hypothetical which is common in

[1] For studies on willingness to compete, see for example Hogarth et al. (2011) and Niederle and Vesterlund (2007) or the literature review by Croson and Gneezy (2009). For studies on different outcomes between the genders in competitive settings, see for example Gneezy et al. (2003) and Gneezy and Rustichini (2004).

[2] See for example Örs et al. (2012), Wozniak (2012), Jurajda and Munich (2011), Paserman (2010), Kleinjans (2009), and Lavy (2008).

experimental studies. Third, since we study actual reforms rather than an experiment, we also avoid problems that can arise if subjects respond to an experimental setting by adjusting their behavior.

In the empirical analysis we define the schools in Oslo and Hordaland as treated in the years after the reforms and counties with the neighborhood rule throughout the period studied as control group. This allows us to estimate the difference-in-differences between the students subject to the reform (treated) and students that are not affected by the reform (non-treated).

Our difference-in-differences approach indicates clear differences between the genders with regards to how student motivation is affected by the reform. Whereas boys' motivation seem to remain unaffected by the reform, there is a sizable negative effect on girls' motivation. Our findings are thus in line with findings in the experimental literature indicating that females are less willing to compete than males. Importantly, our study of two similar reforms in two different counties at two different points in time yield very similar results. We think that this clearly adds credibility to our estimates.

The remainder of the paper is organized as follows: Section 2 provides a brief overview of the related literature. In Section 3 we present the relevant institutional background and discuss the reforms of interest. We present our data and empirical strategy in Section 4, before the main results are reported in Section 5. In Section 6 we present some specification tests, before we summarize our findings in Section 7.

2 Related literature

Gender differences have received much attention in the experimental economics literature. The typical findings are that males are more willing to compete than females, and that males also perform better than females in competitive settings. The experimental data varies from controlled "lab experiments" on adults (e.g. Gneezy et al. (2003) and Niederle and Vesterlund (2007)) or school children (e.g. Gneezy and Rustichini (2004) and Cárdenas et al. (2012)[3]), to the use of data tapped from televised game shows (e.g. Hogarth et al., 2011).

This paper adds to the recent literature using field data. Interestingly, the findings in this literature are more mixed than in the experimental studies. Jurajda and Munich (2011) study multiple university entrance exams taken by the same individuals and find that women do not shy away from applying to more competitive institutions. Interestingly, men do outperform

[3]The study by Càrdenas et al. (2012) is interesting in the sense that their study of Swedish children indicates an unambiguity in competitiveness. Whereas girls seem to improve their performance more than boys when forced to compete, boys are more likely to choose to compete.

women in entrance exams for these more prestigious institutions. There exist, however, no such differences for the exams for the less competitive schools.

Kleinjans (2009) investigates whether gender differences in preferences for competition can explain some of the differences in career choices. Her findings suggest that women's greater distaste for competition decrease educational achievement. It can also explain part of the gender segregation in occupational fields. Specifically, accounting for distaste for competition seems to reduce gender segregation in the fields of law, business and management, health, and education.

Örs et al. (2012) find that males outperform females on high-competitive tests, while it is the other way around in tests with low competition. However, Lavy (2008) finds no gender-based difference in a rank-order tournament among high school teachers where prizes are based on their students' performance. Paserman (2010), examining the performance of highly competitive professional tennis players during grand slam tournaments, finds no difference in forced errors across genders during critical points of the game. In another paper studying professional tennis players, Wozniak (2012) studies the effect from relative performance feedback. His findings suggest that the probability that females choose to enter tournaments depend on their performance in a series of earlier tournaments. The same probability for men, on the other hand, only depends on the performance in the two last tournaments.

Our study of student motivation shares important similarities with the study by Jürges and Schneider (2010). They find that, to quote the title of their paper, "Central exit examinations increase performance... but take the fun out of mathematics". Specifically, using student answers from the TIMSS questionnaire, they find that students in German states with central exit examinations are consistently less likely to enjoy doing mathematics, and to find that mathematics is easy, and more likely to find mathematics boring. Unlike us, however, Jürges and Schneider do not focus on gender differences in their study.

In a broader context, this paper is also related to the school choice literature, which mainly analyzes how parental school choice works as an incentive for the school to improve productivity and quality of schooling.[4] There are two main mechanisms outlined in this literature. First, students in below average schools get the chance to attend better schools, thus improving their achievement. Second, once students start to leave the neighborhood school in favor of a better school, schools will want to compete for the best students and therefore have incentives to provide the best possible education to the student consumer.[5]

[4]Economics of school choice emerge from the work by Friedman (1955). He argued against a neighborhood rule that require parents to send their child to the nearest public school. He argued that this organization of schooling is economically inefficient as it prohibits competition among schools.

[5]There is a large literature analyzing different aspects of school choice. Figlio and Hart (2011) examines

In a study of the Hordaland reform, Haraldsvik (2012) identifies a positive effect on lower secondary student achievement, using a difference-in-differences approach similar to ours. Hence, the incentives implied by the increased competition seem to be of sufficient magnitude to affect student behavior. Whereas gender differences is a central theme in the present paper, Haraldsvik does not discuss this topic in her paper.

Our study of another outcome than test scores can be motivated by a growing literature that distinguishes between cognitive skills (e.g. test scores or exam grades) and non-cognitive skills. Studies by Heckman et al. (2006) and Heckman and Rubinstein (2001) find that non-cognitive skills play a substantial role in determining labor market outcomes and degree attainment. Hence, some worry that studies that focus exclusively on test scores fail to capture some important elements of school production.

3 Institutional background and the reforms

Since 1994, all who graduate from lower secondary education in Norway have a legal right to three years of upper secondary education. The legal right lasts for five years after completed lower secondary education, but expires prior to this if the student reaches 24 years of age.[6] The upper secondary schools are owned and operated by the 19 Norwegian counties, and the admission systems vary from county to county.

In early spring, around the same time as the survey is conducted, the lower secondary graduates rank their upper secondary programs and schools of choice. Upper secondary education is divided into two main tracks, academic and vocational. In a given cohort, almost 50 percent of the students attend the academic track. The academic track is divided in three programs, and most (about 40 percent of the cohort) follow the general academic education program. Because of the low costs associated with offering this program, it is offered at roughly half of the schools. The vocational track consists of nine different education programs, where no program singles out as the dominant one. Due to the limited demand for these programs and the fact that they can be costly to offer, we have that most vocational tracks are offered only at a few schools within each county. We thus have that students applying for these face a limited school choice even under the performance based system.

The fact that students choosing different tracks are not affected by the reform to the same extent can be a potential worry for studies such as ours. It is therefore crucial to note that the share of students that choose the academic track is quite evenly distributed between the

how student test scores are influenced by school competition. This paper also provides an overview of research done on this field.

[6] No students fail any grade in Norwegian compulsory education. The 24 year rule is thus, in practice, a non-binding constraint.

genders. Hence, we do not expect different treatment of boys and girls to be a problem for our study.

The legal right to upper secondary education relate to education program and not to school, so even though the students may prioritize among schools, they are not entitled admission to a specific school. When subject to a neighborhood admission system of upper secondary schooling, the students have limited possibility to influence school admission apart from choosing education program. The neighborhood admission system is quite rigid, compared to the performance based system where the students can prioritize among all upper secondary schools in the county. The schools that have more applicants than study places then rank students based on their final grade point average (GPA) from lower secondary school. The GPA consists of teacher assessed grades and grades from final exit examinations, and all grades are weighted equally. Hence, under the performance based system, which school a student is admitted to is a function of his achievement in lower secondary.

In this paper, we study changes in the admission systems in two counties, Oslo and Hordaland. The reforms in the two counties are quite similar, but were implemented with a few years in between. Hordaland went first, and in the fall of 2004, the county announced that the existing neighborhood school system would be replaced by performance based ranking. The new system was effective from the following school year (2005/06). Hence, the students graduating from lower secondary education in Hordaland in 2005 are the first to be affected by this reform. The reform in Oslo was announced in the fall of 2008. The reform involved that the existing system where 50 percent of the slots at each school was reserved for the students residing in the neighborhood of the school would be replaced by performance based ranking across all schools within the county.[7] The new system was effective from the following school year (2009/10), i.e. the students graduating from lower secondary education in Oslo in 2009 are the first to be affected by this reform. As we see, Oslo did not go from one extreme to another, and was thus not quite as clear-cut as the reform in Hordaland. Hence, one may expect the reform in Hordaland to have stronger effects.

To conduct a difference-in-differences analysis we will compare the development in Oslo and Hordaland with a control group consisting of the 10 Norwegian counties that had a neighborhood rule throughout the period studied.[8]

[7] Prior to the reform in 2009/10, Oslo was divided into four such neighborhoods.

[8] There are 19 counties in Norway. The counties in the control group are Østfold, Hedmark, Buskerud, Telemark, Vest-Agder, Sogn og Fjordane, Sør-Trøndelag, Nord-Trøndelag, Nordland and Troms. Note that there is a minor difference in the control group when studying the two reforms. Two of the local governments in the county of Sør-Trøndelag (Trondheim and Klæbu) are omitted from the reference group when studying Hordaland. This is due to a reform in these local governments within the time period studied. See Haraldsvik (2012) for more details.

4 Data and empirical strategy

4.1 Survey data

Our left hand side variable is tapped from survey data collected by the Norwegian Directorate for Education and Training since the school year 2003/04. The survey is sent out to lower secondary schools once every year, and is answered by the last year students during the period January-April.[9] Among the questions is one that is well suited for our purpose, *"How interested are you in learning at school?"*. The response 1 indicates very low interest, while 5 (4 prior to 2007) indicates that students are highly interested in their schoolwork. We interpret this as a measure of the students' intrinsic motivation. The question we use is similar to the TIMSS questions about student attitudes towards schooling in German states studied by Jürges and Schneider (2010).

Unfortunately, the survey was not compulsory in the first year. In addition, many schools did not report data for girls and boys separately in the second year. The data for 2004 and 2005 thus include relatively few observations. This may introduce selection issues, since it is not necessarily a representative sample of schools that reported data in the first two years. Further, the data for 2006 is missing in the database.[10] Moreover, in 2007 several questions in the survey was dropped and new questions were included. The scale also changed from 1-4 to 1-5. Thereafter, the number of schools in the data set has been very stable, and the design of the survey has remained unchanged. Hence, the data from 2007 and onwards is clearly of a higher quality than for the earlier years.

The timing of the reform in Oslo makes it possible to restrict the analysis to the new version of the survey when studying this reform. We thus use survey data for the five years 2007-2011 in our analysis of the Oslo reform, giving us two years of pre-treatment data and three years of post-treatment data.[11] Since the Hordaland reform took place in 2005, we have to use data also for the first years when studying this reform. Hence, we use all available survey data for the period 2004-2011 (keeping in mind that 2006 is missing) in our study of the Hordaland reform. Since the formulation of the specific question we study is identical in the two surveys, there should not be any fundamental problems with such a crossover between the surveys. The main worry is thus the fact that both the treatment and control

[9] In the remainder we will refer to the year the survey was conducted. E.g. when referring to the school year 2003/04, we write only 2004.

[10] The data for this year is not omitted for any particular reason. In fact, it was simply not included when the directorate updated its web service. According to the directorate this is likely due to technical and/or human errors. They have also informed us that the data may be uploaded "sometime in the future".

[11] We have estimated our main models for Oslo also including the 2005 data. This does not change our main results. The data for 2004 cannot be used in the Oslo analysis due to a change in the admission system with effect from 2004.

group changed due to the large increase in participating schools after the first two years. We will get back to this issue when discussing the empirical findings below. There is also another drawback with the data for Hordaland. Since the reform took place in 2005, the survey in 2004 provides the only data available for the pre-reform period. Hence, we cannot conduct a placebo test of the kind we do for Oslo in Table 3 below. Because of the shortcomings with the data for Hordaland, the results for Oslo are likely to be the most credible. We do think, however, that our study of the reform in Hordaland will be an interesting supplement to the study of the Oslo reform. If studies of two similar reforms, in two different counties at two different points in time yield similar results, this will clearly enable us to make stronger conclusions than if the study was based on only the reform in one of the two counties.

The students' responses to the survey, aggregated to the school level for each year, are available through the directorate's webpage. Due to the aggregation, our left hand side variable is continuous rather than an ordinal scale as it would be on the individual level. In order to secure the anonymity of the individual student, results for schools with less than 10 students in the cohort are not made public. This also applies for the separate male and female scores, where the score values are omitted when there are less than 10 males/females in the cohort.

To secure that the treatment and control groups are consistent across specifications, we only use data for schools where data are published for both genders. As a consequence, all schools with cohorts of less than 20 students are excluded. Some schools with a few more than 20 students in the cohort are also excluded, since the genders in most cases are not evenly distributed. Schools are also omitted from publication if the response rate is lower than 50 percent, regardless of their size. As a consequence we are left with 48 out of a total of 66 schools in the treatment group and 341 out of a total of 606 in the control group when studying the Oslo reform. Likewise the sample of schools is reduced to 73 from a total of 107 in the treatment group and 336 out of 604 in the control group when studying the Hordaland reform.

A noteworthy side-effect by the regulations for the publication of results is that they reduce the number of small schools, making the control and treatment groups more homogenous with regards to school size. This is illustrated in Table 1, which compares the full population of schools to what we have in our sample. The difference in cohort size between the treatment and control groups is reduced with about 29 percent and 24 percent when comparing the population to the sample in Oslo and Hordaland, respectively. The removal of small schools also reduces differences in urbanity to some extent, since the smallest schools most often are found in rural areas.

The fact that small schools in rural areas are omitted reduces the potential worry that

Table 1: Population and sample

	Population		Sample	
	Oslo	Control group	Oslo	Control group
Number of municipalities	1	237	1	183
Number of schools	66	606	48	341
Mean number of student	80.3	47.2	100.3	76.9
Mean number of girls	38.7	22.9	48.9	37.5
Mean number of boys	41.6	24.3	51.4	39.4

	Population		Sample	
	Hordaland	Control group	Hordaland	Control group
Number of municipalities	33	235	26	183
Number of schools	107	604	73	336
Mean number of student	63.4	44.9	88.9	74.8
Mean number of girls	30.8	21.8	43.2	36.4
Mean number of boys	32.6	23.1	45.6	38.4

schools in the urban county Oslo differs from the control group. We expect that the level of urbanity to a large extent will be captured by school fixed effects, which we will use throughout the empirical study. Anyway, we note that Hordaland has a much more diversified population pattern than Oslo. Hordaland includes both a large city,[12] small towns and rural areas.

In order to ease the interpretation of the results, we have standardized the responses into a z-score with zero mean and a standard deviation of unity over the period studied.[13] Since there was a change in the survey's scale (from 1-4 to 1-5) in 2007, this is done by standardizing the residuals when regressing (OLS) the raw score on motivation as the dependent variable and year dummies as independent variables.

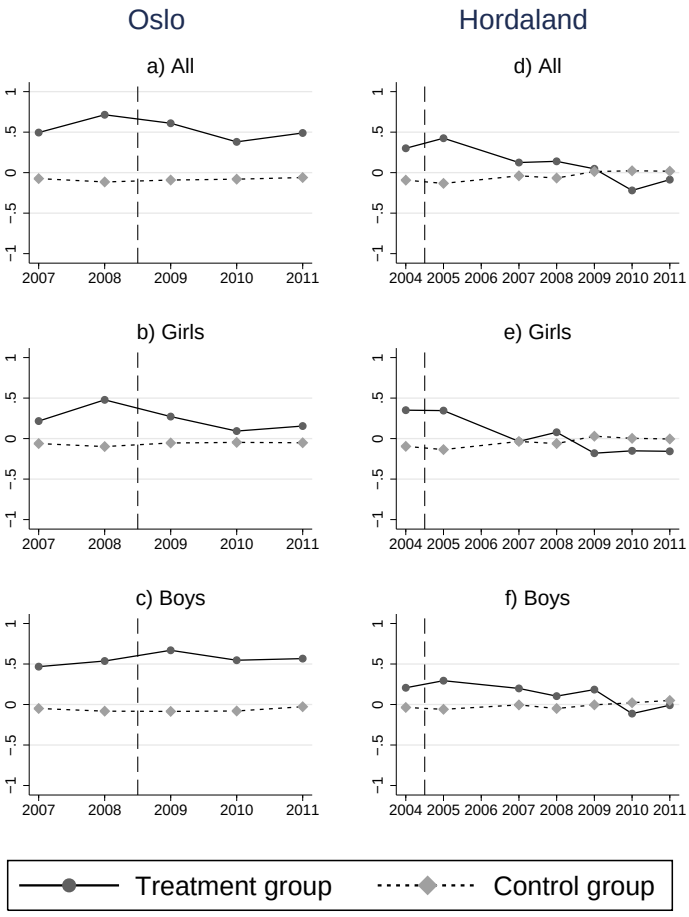
Figure 1 shows the descriptive statistics for the variable of interest in Oslo and the control group, and Hordaland and the control group. The descriptive statistics for Oslo are given in a)-c). We observe that whereas the trend is flat throughout the period for the control group there is some variation in Oslo. The general picture is that students in Oslo had higher reported motivation than the control group the whole period, but the difference decreased after the reform. Both when looking at the full population of students in Figure 1a) and girls in Figure 1b) we observe a slight increase in motivation for Oslo students prior to the reform, followed by a decrease in the period after the reform. The drop is particularly clear when looking at the girls, but less obvious when studying the full pool of students. The reason for this becomes obvious when we look at boys in Figure 1c). We observe that the curve for the boys in Oslo remain almost completely flat throughout the period, giving that all the variation in Figure 1a) is due to changes in the motivation of girls.

Next, we turn to the comparison of Hordaland and the control group, in d)-f). Interest-

[12]The second largest city in Norway, Bergen, is located in Hordaland.

[13]The motivation score is close to being normally distributed in both of our cases. See Appendix Table A for descriptive statistics on the raw score.

Figure 1: Descriptive statistics on motivation



Weighted means using the number of all students, girls and boys, respectively, as weights. Standardized values.

ingly all three graphs show that Hordaland students had higher motivation than the control group prior to the reform, but that the control group caught up with Hordaland after the reform. On first sight, the gender differences are thus less apparent in the Hordaland data than in the Oslo data, even though the change in motivation after the reform does seem to be larger for girls than boys. Further, whereas there seems to be a drop in girls' motivation shortly after the reform, the trend is flat for boys in several years before dropping slightly towards the end of the sample. Hence, we also observe indications of gender differences in the descriptive statistics for the Hordaland reform.

In general, the descriptive statistics in Figure 1 indicate that increased competition has adverse consequences for girls' intrinsic motivation for, while it is less obvious whether or not boys are affected by the reform. We must, however, conduct a more formal econometric investigation before we can draw any real conclusions.

4.2 Econometric specification

In order to study whether the reforms affect intrinsic motivation we estimate various versions of

$$y_{it}^{g} = \beta^{g}\text{Treatment}_{\text{it}} + \mathbf{Controls_{it}'}\boldsymbol{\kappa}^{\mathbf{g}} + \alpha_i^g + \delta_t^g + \epsilon_{it}^g \tag{1}$$

y_{it}^g is the reported motivation in school i in year t and g = (all, boys, girls). When estimating regressions where the genders are pooled, we weight the regressions by the average total cohort size. When estimating using the genders separately, we weight the regressions by the number of students of the specific gender in the cohort.

The variable "Treatment" is the interaction term between treated schools and dummies for each year of the post-treatment period, or in some specification simply a dummy for being in the post-treatment period. α_i and δ_t capture school and time (i.e. year) fixed effects, respectively. Importantly, these capture both differences between schools that are constant over time, and changes from one year to another that are equal for all schools. Hence, the coefficient β captures the difference in differences between schools that are subject to the reform (Oslo/Hordaland) and those that are not (the control group). ϵ_{it} is an error term.

In addition, we include a vector of control variables that may affect student motivation, in order to ensure that our estimates for the treatment effect are not plagued by bias due to omitted variables. Since these should not vary much within school over time, and neither cause nor be caused by the reform, we do not expect them to play an important role in the analysis. Anyway, the vector of control variables include the size of the cohort and a proxy for class size, the share of girls in each cohort, total resources spent on schools in the municipalities during the school year per student and some important socioeconomic characteristics of the local government.[14] The socioeconomic characteristics of the local government include the (log) average private income, unemployment rates for the age groups 16-24 years and 25-66 years and the share of single supporters, immigrants, people with high school and short and long university education. Descriptive statistics for the control variables are reported in Appendix Table B.

[14] The proxy for class size is defined as the number of students at levels 8 to 10 multiplied with the average number of hours in class devided by total teaching hours. Budgets are weighted with 5/12 and 7/12 of the budget for the year of the first and second semester, respectively.

5 Main results

5.1 The Oslo reform

The results from our analysis of the Oslo reform are presented in Table 2. In Columns (A)-(C) we study the effect on girls. We start out in Columns (A) and (B) with a simple formulation where all years of post-treatment are pooled together. The difference between Columns (A) and (B) is that the vector of control variables is omitted in (A), while included in (B). We observe that the coefficient increases slightly in absolute value when including the controls, but that the difference is small compared to the standard error.

Table 2: Estimations of student motivation.

	(A)	(B)	(C)	(D)	(E)	(F)	(G)
VARIABLES	Girls	Girls	Girls	Boys	Boys	Boys	All
Treatment	-0.178	-0.254		0.0967	0.192		
	(0.132)	(0.210)		(0.101)	(0.160)		
Treatment*09			-0.176			0.221	-0.0297
			(0.209)			(0.170)	(0.202)
Treatment*10			-0.516*			0.0973	-0.402
			(0.288)			(0.220)	(0.265)
Treatment*11			-0.525*			0.0219	-0.374
			(0.292)			(0.228)	(0.268)
Control variables	No	Yes	Yes	No	Yes	Yes	Yes
Observations	1,648	1,648	1,648	1,648	1,648	1,648	1,648
No. of schools	389	389	389	389	389	389	389

Robust standard errors (clustered on the school level) in parentheses.
School fixed effects, a constant term and time dummies are included in all regressions.
Weighted regressions using the number of girls/boys/total students as weights.
*** p<0.01, ** p<0.05, * p<0.1

The sign of the treatment effect is negative, as expected when studying girls, but not significantly so in neither Column (A) nor Column (B). In order to investigate this further, Column (C) look at the effect year by year after the implementation of the reform. We observe that the effect is insignificant in the first year, but increases sharply and becomes significant at the 10 percent level in the following two years. The coefficients for the treatment effect in 2010 and 2011 are sizable, indicating that the intrinsic motivation is roughly 52-53 percent of a standard deviation lower two and three years after the reform implementation. The effect thus seems quite strong, but one should note that the standard deviation of the motivation variable is not very large. A coefficient of -0.52 is equivalent to a 0.11 reduction in the survey scale points, which again translate to 3 percent of the average value.[15]

The observation that we do not find any immediate effects is not particularly surprising. The reform was announced shortly before the survey was conducted for the first treated

[15] Calculated from the means and standard deviations reported in Appendix Table A. The results are not sensitive to the use of weights. If we do not use weights when regressing girls' motivation, the coefficients (std. errors) are -0.536 (0.300), and -0.627 (0.300) in 2010 and 2011, respectively. The results without weights are also similar as in the weighted regressions when studying boys and the genders pooled.

cohort. Hence, they had very little time to understand the implications of the system change. Moreover, it is likely that the students in the first treated cohort would have had a limited understanding of the full implications of the reform anyway. The reason is that they had very little (if any) grounds for comparison between the two regimes, since they were the first to be affected by it.

When we look at boys in Columns (D)-(F) the picture is very different. The coefficients for the treatment consistently come out as positive, on face value indicating an opposite effect on boys compared to girls. The coefficients are, however, far from being significant, suggesting a zero effect on the motivation of boys from the reform.

Interestingly, we have that the coefficients for boys and girls are significantly different both when using the single treatment dummy specification, and when analyzing the treatment effect year by year.[16] Hence, despite the fact that the treatment is not significantly negative for girls per se in the one dummy formulation and in 2009, girls' motivation is still significantly reduced compared to that of boys.

Finally, we in Column (G) pool boys and girls together, to see if there are any total effects. We observe that while the coefficients are consistently negative, they are not close to being significant at any conventional levels of significance. The main take-away from Table 2 is that it confirms the impression from the descriptive statistics reported in Figure 1a)-c). There seems to be a clear negative effect on the motivation of girls in Oslo from the reform, but no effect on the motivation of boys. Our findings so far thus seem to be clearly in line with the earlier studies concluding that females are less willing to compete than males. See e.g. Hogarth et al. (2011) and Niederle and Vesterlund (2007) or the literature review by Croson and Gneezy (2009).

There is, however, one concern that must be addressed before firm conclusions can be drawn. It is possible that the estimated reform effect reflects time trends rather than a true treatment effect. In order to address this concern, we investigate lower secondary performance trends both before and after the reform. This is done by estimating models where the treatment is introduced prior to the reform, known as placebo difference in difference. From Figure 1 we can clearly see that the coefficients for Oslo do not capture a lasting negative trend starting already prior to the reform, since if anything, there seems to be a weak positive trend in Oslo prior to the reform.

If our test rejects a common pre-treatment trend, this can make the interpretation of our estimates difficult for two reasons. First, if the reform leads to a shift from a positive to a negative trend, our estimates will in fact underestimate the true effect. Second, if the

[16] The test involves estimating a model including the same variables as in the models with control variables in addition to gender specific school effects, and interaction terms between a gender dummy and all other variables.

Table 3: Placebo tests.

VARIABLES	(A) Girls	(B) Boys	(C) All
Treatment*08	0.291	0.0382	0.212
	(0.199)	(0.185)	(0.181)
Treatment*09	0.0300	0.248	0.121
	(0.237)	(0.227)	(0.240)
Treatment*10	-0.306	0.125	-0.249
	(0.311)	(0.258)	(0.292)
Treatment*11	-0.276	0.0546	-0.193
	(0.344)	(0.287)	(0.319)
Observations	1,648	1,648	1,648
No. of schools	389	389	389

Robust standard errors (clustered on the school level) in parentheses. School fixed effects, a constant term, control variables and time dummies are included in all regressions. Weighted regressions using the number of girls/boys/total students as weights.
*** p<0.01, ** p<0.05, * p<0.1

positive trend is rather followed by a reversion towards a common trend, independent of the reform, our estimates will overestimate the treatment effect. The results from the placebo test are reported in Table 3. The reform effects are now estimated relative to 2007. As we see, there are no signs of such a trend. The coefficient for 2008 (the placebo year) comes out as positive as expected from the graphical illustration, but is insignificant. We can thus not reject that the pre-treatment trends are identical.

5.2 The Hordaland reform

The results from the analysis using the Hordaland reform is reported in Table 4. The year prior to the announcement of the reform (2004) acts as the pre-treatment comparison. The table has the same structure as Table 2. When looking at the girls in Columns (A)-(C) the first thing we note is that the effect seems to be stronger, both in terms of coefficient size and statistical significance than for girls in Oslo. In particular, the coefficients for the treatment effect are significantly negative also when pooling all post-treatment together in Columns (A) and (B).

From Column (C) we observe that, as in Oslo, the treatment effect comes out as insignificantly negative in the first year. The effect three years after the reform (2007) is very similar to the effect we find 2-3 years after the reform in Oslo (2010 and 2011), about 53 percent of a standard deviation. The significance is stronger though, since the coefficient is significant even at the 5 percent level in Hordaland, while only at the 10 percent level in Oslo. Interestingly, the effect is a bit weaker, and falls short of significance in 2008 year, before the coefficients increase in absolute value in the remaining years of the sample.

In the three last years, the treatment effect ranges from about 60-70 percent of a standard

Table 4: Estimations of student motivation using data from Hordaland.

VARIABLES	(A) Girls	(B) Girls	(C) Girls	(D) Boys	(E) Boys	(F) Boys	(G) All
Treatment	-0.491***	-0.395**		-0.0987	-0.0477		
	(0.181)	(0.195)		(0.191)	(0.205)		
Treatment*05			-0.0502			0.0966	0.104
			(0.260)			(0.248)	(0.260)
Treatment*07			-0.526**			0.00157	-0.273
			(0.231)			(0.268)	(0.267)
Treatment*08			-0.332			-0.0854	-0.229
			(0.224)			(0.224)	(0.232)
Treatment*09			-0.641***			-0.0100	-0.342
			(0.227)			(0.233)	(0.251)
Treatment*10			-0.594***			-0.304	-0.622***
			(0.198)			(0.234)	(0.231)
Treatment*11			-0.698***			-0.329	-0.607**
			(0.217)			(0.243)	(0.252)
Control variables	No	Yes	Yes	No	Yes	Yes	Yes
Observations	2,017	2,017	2,017	2,017	2,017	2,017	2,017
No. of schools	409	409	409	409	409	409	409

Robust standard errors (clustered on the school level) in parentheses.
School fixed effects, a constant term and time dummies are included in all regressions.
Weighted regressions using the number of girls/boys/total students as weights.
*** p<0.01, ** p<0.05, * p<0.1

deviation, and is significant even at the 1 percent level in all three years. This is equivalent to about a 0.15 reduction in the survey scale points after the change of the scaling in 2007 (see Appendix Table A).[17] This translates into 3-4 percent of the average value in this period.

As in the study of the Oslo reform, the picture is very different when studying the boys in Columns (D)-(F). Even though the coefficients for the treatment effect mostly come out as negative, as for girls, they are never close to being significant at any conventional level of significance. This is also reflected in Column (G) where the genders are pooled, where only the two last years come out as significant.

Despite the clear significance for girls, and the insignificant estimates for boys, the test for different effects between the genders actually provide slightly less clear-cut conclusions than when studying the Oslo reform. In the one-dummy formulation the t-test for the difference is about 1.42 in absolute value. However, the difference is significant in the years 2007 and 2009. The reason is likely that there are some weak indications in Table 4 that boys in Hordaland also experience a reduction in motivation following the reform, as we also noticed when studying Figure 1f). One possible reason for this is that the reform in Hordaland was more profound than the reform in Oslo (see discussion in Section 3).

Still, the main take-away from Table 4 is that it mirrors the observations from Figure 1e)

[17] Again, the results are not sensitive to the use of weights. If we do not use weights when regressing girls' motivation, the coefficients (std. errors) are -0.599 (0.235), -0.490 (0.223), -0.791 (0.242), -0.696 (0.213), and -0.767 (0.215) in 2007 to 2011, respectively. The results without weights are also similar as in the weighted regressions when studying boys and the genders pooled.

since the motivation of girls is reduced following the reform. Further, while the sign of the treatment effect for boys is negative, reflecting the small negative trend in Figure 1f), there is no significant treatment effect. The results for Hordaland thus support the findings from the study of the Oslo reform in Table 2. We think that the fact that our study of two similar reforms in two different counties at two different points in time yield very similar results clearly adds credibility to our findings.

As mentioned in the data discussion in Section 4, the survey was not compulsory the first year. This results in relatively few observations in 2004, and also leads to a worry about potential sorting since it is not necessarily a representative sample of schools that chose to conduct the survey this year. To address this, we have also conducted difference-in-differences regressions restricted only to schools that participate in the survey that year. The findings are reported in Appendix C. We observe that, as in the main results, boys do not respond to the treatment, whereas the introduction of increased competition has adverse consequences for girls' motivation. In fact, the reported coefficients are stronger for girls in the last period compared to the analyses of the full sample, but the effect in 2007 is no longer statistically significant.

Unfortunately we can, due to the reasons discussed in Section 4, not test whether the coefficients for the Hordaland reform is a treatment effect or simply captures some time trend. It should be noted, however, that the clear rejection of such trends in Oslo at least indicates that such trends may not be a serious problem for our study.

6 Specification tests

The results above indicate gender differences in competitiveness. However, before any certain conclusions can be drawn, we must address a few concerns. The first is regarding a difference in the control groups when splitting the genders. When we study boys (girls) in our benchmark analyses, the control group also consists only of boys (girls). One may thus worry that the difference in the results between the genders may be driven by differences in the control groups. In order to investigate this, we have estimated the effect on boys and girls using the full sample of students in the control group. The results are very similar to those reported, and differences in the control group are thus not driving the difference in the results for boys and girls.[18]

Some suggest that heterogeneity may also follow other dimensions. Donze and Gunnes

[18]The reform effects for girls in 2010 and 2011 are -0.45 and -0.55, respectively, when studying the Oslo reform. When studying the reform in Hordaland the corresponding effects are about -0.54, -0.38, -0.52, -0.57 and -0.80 for the years 2007-2011, respectively.

(2011) argue, equivalent to the findings in Leuven et al. (2010), that while a competitive environment may motivate high-ability students, it can have adverse consequences for low-ability or risk averse students. Because we operate with school averages we cannot do a proper analysis of this question. We have, however, split the schools into different categories based on their achievement when estimating the model.[19] We find no clear-cut patterns, but the gender differences still occur in the split samples, indicating that our benchmark results are not invalidated by heterogeneity on this alternative dimension.

Table 5: Robustness test: Excluding one county from the control group at the time when studying the Oslo reform.

County excluded	Østfold		Hedmark		Buskerud		Telemark		Vest-Agder	
	Girls	Boys	Girls	Boys	Girls	Boys	Girls	Boys	Girls	Boys
Treatment*09	-0.152	0.212	-0.171	0.208	-0.195	0.122	-0.196	0.224	-0.128	0.316*
	(0.217)	(0.174)	(0.214)	(0.172)	(0.217)	(0.172)	(0.212)	(0.172)	(0.213)	(0.172)
Treatment*10	-0.480	0.027	-0.443	0.182	-0.598**	-0.037	-0.534*	0.083	-0.516*	0.217
	(0.296)	(0.230)	(0.294)	(0.220)	(0.301)	(0.227)	(0.290)	(0.221)	(0.297)	(0.215)
Treatment*11	-0.468	-0.003	-0.537*	0.059	-0.547*	-0.141	-0.547*	0.035	-0.479	0.192
	(0.301)	(0.239)	(0.301)	(0.233)	(0.305)	(0.231)	(0.294)	(0.229)	(0.304)	(0.225)
Observations	1,467	1,467	1,518	1,518	1,469	1,469	1,510	1,510	1,510	1,510
No. of schools	351	351	359	359	346	346	354	354	359	359

County excluded	Sogn og Fjordane		Sør-Trøndelag		Nord-Trøndelag		Nordland		Troms	
	Girls	Boys	Girls	Boys	Girls	Boys	Girls	Boys	Girls	Boys
Treatment*09	-0.203	0.215	-0.354	0.265	-0.180	0.197	-0.173	0.228	-0.129	0.221
	(0.214)	(0.173)	(0.238)	(0.193)	(0.213)	(0.171)	(0.214)	(0.173)	(0.211)	(0.177)
Treatment*10	-0.491*	0.118	-0.782**	0.135	-0.504*	0.095	-0.562*	0.094	-0.413	0.063
	(0.294)	(0.225)	(0.329)	(0.260)	(0.292)	(0.222)	(0.300)	(0.233)	(0.294)	(0.239)
Treatment*11	-0.544*	0.023	-0.862**	0.032	-0.525*	0.003	-0.526*	0.027	-0.416	-0.001
	(0.302)	(0.234)	(0.345)	(0.271)	(0.297)	(0.230)	(0.301)	(0.238)	(0.297)	(0.249)
Observations	1,545	1,545	1,454	1,454	1,553	1,553	1,486	1,486	1,553	1,553
No. of schools	360	360	347	347	366	366	342	342	365	365

Robust standard errors (clustered on the school level) in parentheses.
School fixed effects, a constant term, control variables and time dummies are included in all regressions.
Weighted regressions using the number of girls/boys as weights.
*** p<0.01, ** p<0.05, * p<0.1

Next, we acknowledge that the validity of our diff-in-diff analysis relies on the choice of a proper control group. In our main analyses, we have included all counties that had a neighborhood system throughout the period studied. The reasoning behind this is to compare admission systems that were similar prior to the reforms in Oslo and Hordaland. The Norwegian counties are, however, quite heterogeneous with respect to geography, population size and demography. A potential worry is thus that the results can be driven by such differences.

To test the robustness of our main results, we have excluded one county at the time from the control group. In order to save space, we only report the results for the genders separately. This gives us a total of 20 new regressions for each of the treated counties. The results when studying the reforms in Oslo and Hordaland are reported in Tables 5 and 6, respectively. We observe that the results are quite robust, and thus it seems that our findings are not driven by any particular counties. We do, however, note that the results for Oslo

[19] We split the school based on average performance on exit examinations in the years prior to the reforms, since post-reform achievement is endogenous.

Table 6: Robustness test: Excluding one county from the control group at the time when studying the Hordaland reform.

County excluded	Østfold		Hedmark		Buskerud		Telemark		Vest-Agder	
	Girls	Boys	Girls	Boys	Girls	Boys	Girls	Boys	Girls	Boys
Treatment*05	-0.063	0.126	-0.055	0.180	-0.003	0.144	-0.054	0.068	-0.040	0.101
	(0.266)	(0.257)	(0.263)	(0.249)	(0.266)	(0.259)	(0.263)	(0.251)	(0.267)	(0.248)
Treatment*07	-0.536**	0.008	-0.558**	0.061	-0.535**	0.019	-0.526**	-0.003	-0.462*	0.077
	(0.238)	(0.276)	(0.234)	(0.266)	(0.235)	(0.277)	(0.235)	(0.273)	(0.239)	(0.269)
Treatment*08	-0.330	-0.037	-0.360	-0.005	-0.318	-0.070	-0.348	-0.111	-0.269	-0.070
	(0.232)	(0.232)	(0.228)	(0.225)	(0.230)	(0.234)	(0.229)	(0.229)	(0.228)	(0.227)
Treatment*09	-0.662***	0.010	-0.626***	0.058	-0.638***	-0.029	-0.648***	-0.006	-0.552**	0.074
	(0.235)	(0.241)	(0.230)	(0.233)	(0.231)	(0.241)	(0.233)	(0.237)	(0.234)	(0.238)
Treatment*10	-0.593***	-0.312	-0.582***	-0.212	-0.637***	-0.302	-0.612***	-0.334	-0.520**	-0.217
	(0.202)	(0.241)	(0.200)	(0.234)	(0.203)	(0.242)	(0.203)	(0.238)	(0.204)	(0.237)
Treatment*11	-0.678***	-0.312	-0.717***	-0.253	-0.667***	-0.349	-0.690***	-0.330	-0.609***	-0.212
	(0.222)	(0.251)	(0.219)	(0.244)	(0.222)	(0.253)	(0.222)	(0.246)	(0.226)	(0.244)
Observations	1,786	1,786	1,846	1,846	1,787	1,787	1,851	1,851	1,847	1,847
No. of schools	369	369	378	378	364	364	373	373	378	378

County excluded	Sogn og Fjordane		Sør-Trøndelag		Nord-Trøndelag		Nordland		Troms	
	Girls	Boys	Girls	Boys	Girls	Boys	Girls	Boys	Girls	Boys
Treatment*05	-0.071	0.056	-0.067	0.104	-0.025	0.052	-0.059	0.056	-0.067	0.080
	(0.257)	(0.249)	(0.261)	(0.250)	(0.258)	(0.247)	(0.263)	(0.249)	(0.257)	(0.249)
Treatment*07	-0.497**	-0.004	-0.549**	0.002	-0.543**	-0.076	-0.537**	-0.058	-0.516**	-0.004
	(0.231)	(0.270)	(0.233)	(0.271)	(0.234)	(0.268)	(0.235)	(0.272)	(0.229)	(0.268)
Treatment*08	-0.279	-0.053	-0.350	-0.096	-0.358	-0.148	-0.338	-0.139	-0.359	-0.119
	(0.223)	(0.227)	(0.226)	(0.228)	(0.227)	(0.224)	(0.230)	(0.228)	(0.221)	(0.223)
Treatment*09	-0.621***	-0.018	-0.680***	-0.010	-0.664***	-0.119	-0.616***	-0.042	-0.651***	-0.005
	(0.226)	(0.236)	(0.228)	(0.238)	(0.230)	(0.231)	(0.232)	(0.237)	(0.224)	(0.231)
Treatment*10	-0.553***	-0.265	-0.616***	-0.339	-0.607***	-0.365	-0.601***	-0.315	-0.579***	-0.360
	(0.198)	(0.237)	(0.200)	(0.239)	(0.200)	(0.234)	(0.201)	(0.237)	(0.197)	(0.237)
Treatment*11	-0.688***	-0.321	-0.757***	-0.370	-0.723***	-0.413*	-0.683***	-0.382	-0.717***	-0.339
	(0.217)	(0.246)	(0.220)	(0.250)	(0.220)	(0.241)	(0.221)	(0.245)	(0.206)	(0.246)
Observations	1,888	1,888	1,884	1,884	1,895	1,895	1,827	1,827	1,900	1,900
No. of schools	380	380	384	384	384	384	360	360	384	384

Robust standard errors (clustered on the school level) in parentheses.
School fixed effects, a constant term, control variables and time dummies are included in all regressions.
Weighted regressions using the number of girls/boys as weights.
*** p<0.01, ** p<0.05, * p<0.1

are a little bit more sensitive than the results for Hordaland. In the cases where Østfold and Troms are excluded in Table 5, the coefficients for the treatment effect on girls fall short of significance. We also note that we actually observe a significantly positive treatment effect on the motivation for boys in the first year when excluding Vest-Agder in the analysis of the Oslo reform.

A final issue is related to the survey data from the Norwegian Directorate for Education and Training. When working with survey data, it is always a worry that the subjects' responses do not necessarily measure exactly what they are meant to do. This could be the case if students who are unsatisfied (satisfied), simply give a poor (good) score on all questions without really considering them separately, or if the students misunderstand the questions. This will create additional noise in our data, but may, in worst case, also lead to estimation of spurious relationships.

In order to investigate this more closely, one may test if the model predicts a treatment effect on variables which the reform likely should not affect. Specifically, we have estimated the model using questions regarding the satisfaction with the physical work environment as dependent variable instead of student motivation. In the new version of the survey, there are 10 such questions and all of these are used when studying Oslo. Due to the changes made

Table 7: Estimations of satisfaction with physical work environment. Oslo reform.

	(A) Indoor air	(B) Indoor temperature	(C) Classrooms	(D) Teaching equipment	(E) School library	(F) Restrooms	(G) Wardrobes & showers	(H) School buildings	(I) Cleaning	(J) Outdoor area
Panel A: Girls										
Treatment*09	0.0215	0.195	0.0436	0.444**	0.168	0.0159	-0.00383	-4.77e-05	0.204	-0.174
	(0.182)	(0.193)	(0.165)	(0.181)	(0.135)	(0.151)	(0.157)	(0.137)	(0.161)	(0.165)
Treatment*10	-0.144	0.0972	0.110	0.0271	0.117	0.0655	0.0455	-0.0418	0.155	-0.159
	(0.224)	(0.244)	(0.229)	(0.224)	(0.189)	(0.230)	(0.244)	(0.204)	(0.215)	(0.219)
Treatment*11	-0.202	0.0376	-0.104	-0.299	0.0193	0.159	-0.0417	-0.254	0.111	-0.231
	(0.264)	(0.267)	(0.250)	(0.241)	(0.212)	(0.286)	(0.273)	(0.236)	(0.238)	(0.249)
Observations	1,623	1,624	1,629	1,629	1,629	1,622	1,615	1,627	1,629	1,630
No. of schools	387	386	387	387	387	387	385	387	387	387
Panel B: Boys										
Treatment*09	-0.115	-0.107	-0.00777	0.366*	0.214	0.00177	-0.00635	0.0190	0.0522	-0.338*
	(0.179)	(0.187)	(0.171)	(0.212)	(0.152)	(0.138)	(0.164)	(0.148)	(0.192)	(0.184)
Treatment*10	-0.143	-0.284	0.103	0.00402	0.157	-0.00499	0.0337	0.0265	-0.0005	-0.271
	(0.213)	(0.221)	(0.188)	(0.227)	(0.192)	(0.195)	(0.233)	(0.203)	(0.228)	(0.223)
Treatment*11	-0.146	-0.293	-0.0753	-0.227	0.0800	0.160	0.0223	-0.0997	0.0067	-0.448*
	(0.253)	(0.251)	(0.226)	(0.246)	(0.208)	(0.241)	(0.257)	(0.226)	(0.236)	(0.251)
Observations	1,626	1,630	1,629	1,630	1,628	1,623	1,627	1,630	1,630	1,630
No. of schools	387	387	387	387	387	386	387	387	387	387

Robust standard errors (clustered on the school level) in parentheses.
School fixed effects, a constant term, control variables and time dummies (not reported) included.
Weighted regressions using the number of girls/boys as weights.
*** p<0.01, ** p<0.05, * p<0.1

to the survey from 2007, only 6 questions on the physical work environment are identical throughout the period studied for Hordaland.

Since the physical work environment should not be affected by the reform, a significantly negative result for these could give reason for caution, since it may indicate that our estimates for motivation captures a broader dissatisfaction. Again, we only report the results from regressions for the genders separate.

The results for Oslo and Hordaland are reported in Tables 7 and 8, respectively. Importantly, the coefficients are mostly far from being significant. Despite a few exceptions, the general impression is that the reported satisfaction with the physical work environment is, as expected, not related to the reform. These results increase our confidence in the survey's ability to measure what it intends to measure, since we do not find that the reform affects variables that it most likely should not affect. Further, the signs are varying, some being negative and some being positive. Positive signs indicate that the negative link between the reform and motivation is not driven by a "general dissatisfaction". Moreover, it is hard to identify any clear gender differences in the tables. Hence, our conclusions about gender differences from the discussion of our main results should not be undermined.[20]

[20] Our procedure is similar to the testing procedure suggested by Carlsen and Johansen (2004). They include a question from the survey that should not be affected by the explanatory variable of interest as a control variable capturing the "general satisfaction". We have also tried this approach. We apply the average satisfaction across the different facility measures as a proxy for general satisfaction. The results show that that satisfaction with facilities is strongly correlated with motivation, even when including school fixed effects, and more importantly that the coefficient for the treatment effect remain unchanged. For Oslo when studying girls the coefficients (std. errors) for treatment are -0.231 (0.205), -0.552 (0.273) and -0.517 (0.279). The treatment effects for boys in Oslo and both genders in the Hordaland study are also approximately the same as in Tables 2 and 4. This is consistent with the findings in the reported test, indicating that there is

Table 8: Estimations of satisfaction with physical work environment. Hordaland reform.

	(A) Classrooms	(B) Teaching equipment	(C) School library	(D) School buildings	(E) Cleaning	(F) Outdoor area
Panel A: Girls						
Treatment*05	0.199	-0.143	0.438	0.0330	-0.0409	0.0674
	(0.158)	(0.151)	(0.332)	(0.147)	(0.179)	(0.171)
Treatment*07	-0.00647	-0.159	0.364	-0.0664	-0.0713	-0.270
	(0.218)	(0.176)	(0.339)	(0.203)	(0.193)	(0.173)
Treatment*08	-0.0913	-0.312*	0.508	-0.114	-0.136	-0.310
	(0.223)	(0.187)	(0.403)	(0.220)	(0.207)	(0.193)
Treatment*09	0.161	-0.116	0.673*	-0.0804	0.108	-0.0927
	(0.265)	(0.216)	(0.381)	(0.243)	(0.228)	(0.198)
Treatment*10	-0.110	-0.206	0.588	-0.210	-0.0970	-0.444**
	(0.251)	(0.206)	(0.365)	(0.242)	(0.235)	(0.213)
Treatment*11	-0.0706	-0.347	0.604	-0.268	-0.0400	-0.211
	(0.262)	(0.233)	(0.369)	(0.243)	(0.243)	(0.209)
Observations	1,987	1,987	1,987	1,984	1,987	1,988
No. of schools	401	401	401	401	401	401
Panel B: Boys						
Treatment*05	0.344**	0.0203	0.611	0.219	0.0111	0.179
	(0.173)	(0.178)	(0.390)	(0.152)	(0.186)	(0.174)
Treatment*07	0.165	-0.0142	0.609	0.0874	0.0135	-0.0015
	(0.196)	(0.175)	(0.386)	(0.189)	(0.200)	(0.171)
Treatment*08	0.0855	-0.144	0.666	-0.0139	-0.115	-0.0463
	(0.211)	(0.218)	(0.454)	(0.211)	(0.242)	(0.181)
Treatment*09	0.158	-0.0989	0.731*	0.0470	-0.0627	0.0370
	(0.230)	(0.220)	(0.411)	(0.214)	(0.238)	(0.185)
Treatment*10	-0.188	-0.265	0.605	-0.171	-0.0851	-0.381*
	(0.221)	(0.214)	(0.427)	(0.206)	(0.243)	(0.209)
Treatment*11	0.102	-0.210	0.666	-0.110	-0.0579	-0.192
	(0.217)	(0.209)	(0.427)	(0.209)	(0.242)	(0.220)
Observations	1,987	1,989	1,986	1,988	1,988	1,988
No. of schools	401	401	401	401	401	401

Robust standard errors (clustered on the school level) in parentheses.
School fixed effects, a constant term, control variables and time dummies
(not reported) included. Weighted regressions using the number of girls/boys as weights.
*** p<0.01, ** p<0.05, * p<0.1

7 Concluding remarks

This paper studies gender differences in competitiveness using data from upper secondary admission reforms in two Norwegian counties and survey data on student motivation. In 2005 and 2009 the counties Hordaland and Oslo, respectively, changed their admission system in upper secondary education from a neighborhood admission system to a system based on performance in lower secondary education (school choice). This introduces high stake competition for students in lower secondary education. Importantly, the fact that several other counties stuck to the neighborhood rule, enables us to study the reform within a difference-in-differences framework.

Our findings suggest that there are clear differences between the genders. Whereas the intrinsic motivation for learning for boys seem to be unaffected by the reform, we identify a sizable negative effect on girls. The fact that our study of two similar reforms in two different counties at two different points in time yield very similar results clearly adds credibility to our findings. Hence, our field study approach produces similar findings as the experimental literature, finding that males in general are more willing to compete than females.

These findings suggest that policy makers face a trade-off between increased performance

no relationship between the reform and satisfaction with facilities.

and decreased motivation, when considering such reforms. It is, however, difficult to give clear-cut policy recommendations, since the magnitude of the effect from reduced motivation on long-run outcomes is uncertain and needs to be studied further.

References

Cárdenas J.-C., Dreber, A., von Essen, E., and Ranehill, E. 2012. "Gender differences in competitiveness and risk taking: Comparing children in Colombia and Sweden" *Journal of Economic Behavior & Organization*, 83, 11-23

Carlsen, F., Johansen, K. 2004. "Subjective measures of employment opportunities and interregional migration" *Labour*, 18, 563-589

Croson, Rachel, Gneezy, Uri. 2009. "Gender Differences in Preferences" *Journal of Economic Literature*, 47 448-474

Donze, Jocelyn, Gunnes, Trude. 2011. "Should economists listen to educational psychologists? Some economics of student motivation" MPRA Paper 31059, University Library of Munich

Figlio, D. and C. Hart: Competitive Effects of Means-Tested School Vouchers, Unpublished manuscript, 2011

Friedman, Milton (1955): The Role of Government in Education, in Robert A. Solow (ed.) Economics and the Public Interest, 123-44, New Brunswick, N.J.: Rutgers University Press.

Gneezy, Uri, Niederle, Muriel, Rustichini, Aldo. 2003. "Performance in Competitive Environments: Gender Differences" *Quarterly Journal of Economics*, 118, 1049-1074

Gneezy, Uri, Rustichini, Aldo. 2004. "Gender and Competition at a Young Age" *American Economic Review (Papers and Proceedings)*, 94, 377-381

Haraldsvik, M. 2012. "Does Performance-Based School Choice Affect Student Achievement?" chapter 3 in Influences on educational outcomes : three essays on the role of parents, peers and choice, Doctoral thesis at NTNU, 2012:346

Heckman, James. J., Rubinstein, Yona. 2001. "The Importance of Noncognitive Skills: Lessons from the GED Testing Program" *American Economic Review: Papers and Proceedings*, 91, 145-149

Heckman, James, J., Stixrud, Jora, Urzua, Sergio. 2006. "The effects of Cognitive and Noncognitive Abilities on Labor Market Outcomes and Social Behavior" *Journal of Labor Economics*, 24, 411-482

Hogarth, Robin M., Karelaia, Natalia, Trujillo, Carlos Andres. 2011. "When should I quit? Gender differences in exiting competitions" *Journal of Economic Behavior & Organization*, 83, 136-150

Jurajda, Stepan, and Daniel Munich. 2011. "Gender gap in admission performance under competitive pressure" *American Economic Review Papers and Proceedings* 101, no. 3:514-518

Jürges, Hendrik, Schneider, Kerstin. 2010. "Central exit examinations increase performance... but take the fun out of mathematics" *Journal of Population Economics*, 23 497-517

Kleinjans, Kristin. 2009. "Do gender differences in preferences for competition matter for occupational expectations?" *Journal of Economic Psychology* 30, no. 5:701–710

Lavy, Victor. 2008. "Gender differences in market competitiveness in a real workplace: Evidence from performance-based pay tournaments among teachers" Working paper no. 14338, National Bureau of Economic Research, Cambridge, MA

Leuven, Edwin, Oosterbeek, Hessel, van der Klaauq, Bas. 2010. "The effect of financial rewards on students' achievement: Evidence from a randomized experiment" *Journal of the European Economic Association*, 8 1243-1265

Niederle, Muriel, Vesterlund, Lise. 2007. "Do women shy away from competition? Do men compete too much?" *Quarterly Journal of Economics*, 122 1067-1101

Örs, E., Palomino, F., Peyrache, E. 2012. "Performance Gender-Gap: Does Competition Matter?" *Journal of Labor Economics*, Forthcoming

Paserman, M. Daniele. 2010. "Gender differences in performance in competitive environ-

ments? Evidence from professional tennis players" Working paper, Boston University, Boston, MA.

Wozniak, D. 2012. "Gender differences in a market with relative performance feedback: Professional tennis players" *Journal of Economic Behavior & Organization*, 83, 158-171

Appendix

A Descriptive statistics, motivation

	2004-2005 (old survey)			2007-2011 (new survey)		
Sample	Treatment	Control	Full	Treatment	Control	Full
Oslo reform						
Girls	-	-	-	4.17	4.10	4.11
				(0.19)	(0.21)	(0.21)
Boys	-	-	-	4.07	3.90	3.93
				(0.21)	(0.24)	(0.24)
All	-	-	-	4.12	4.00	4.02
				(0.17)	(0.18)	(0.19)
Hordaland reform						
Girls	3.45	3.34	3.36	4.09	4.10	4.10
	(0.17)	(0.20)	(0.20)	(0.23)	(0.24)	(0.24)
Boys	3.23	3.16	3.17	3.92	3.90	3.90
	(0.18)	(0.23)	(0.22)	(0.23)	(0.24)	(0.24)
All	3.34	3.25	3.26	4.00	4.00	4.00
	(0.13)	(0.17)	(0.17)	(0.17)	(0.19)	(0.18)

Mean values and standard deviations (in parentheses)
are weighted with the number of girls/boys/total students.
Data for 2006 is missing in the directorate's database.

B Descriptive statistics, control variables

	Oslo study		Hordaland study	
VARIABLES	Oslo	Control group	Hordaland	Control group
Share of girls	0.49	0.49	0.49	0.49
	(0.05)	(0.06)	(0.06)	(0.06)
Class size	17.79	15.63	16.44	15.59
	(2.78)	(2.42)	(2.90)	(13.04)
Municipal spending per student (1,000 NOK)	47.20	50.00	48.50	49.54
	(1.93)	(5.45)	(6.54)	(5.80)
Log (mean gross income)	12.85	12.64	12.67	12.57
	(0.06)	(0.10)	(0.13)	(0.13)
Unemployment rate, 16-24, percent	1.60	2.04	2.10	2.42
	(0.37)	(0.81)	(0.96)	(1.06)
Unemployment rate, 25-66, percent	2.34	1.75	1.97	2.00
	(0.49)	(0.55)	(0.73)	(0.78)
Single supporters (index)	1.78	2.20	2.33	2.50
	(0.21)	(0.51)	(0.60)	(0.71)
Share of immigrants, 0-16, percent	30.81	7.77	6.11	7.12
	(0.73)	(4.80)	(2.84)	(4.90)
Education, share with high school, percent	34.23	44.38	44.50	45.03
	(0.67)	(3.42)	(4.39)	(3.25)
Education, share with BSc, percent	28.48	18.69	20.29	17.59
	(0.40)	(3.72)	(4.07)	(3.56)
Education, share with MSc or PhD, percent	14.59	4.79	6.18	3.95
	(0.86)	(2.79)	(3.29)	(1.90)

Mean values and standard deviations (in parentheses) are weighted with the number of students

C Estimations of student motivation using data from Hordaland

VARIABLES	(A) Girls	(B) Girls	(C) Girls	(D) Boys	(E) Boys	(F) Boys	(G) All
Treatment	-0.492***	-0.493**		-0.0995	-0.0750		
	(0.181)	(0.229)		(0.192)	(0.232)		
Treatment*05			-0.0917			0.0182	-0.0256
			(0.398)			(0.343)	(0.397)
Treatment*07			-0.449			0.00967	-0.160
			(0.358)			(0.404)	(0.418)
Treatment*08			-0.623**			-0.117	-0.420
			(0.304)			(0.260)	(0.259)
Treatment*09			-0.563*			-0.00746	-0.275
			(0.298)			(0.273)	(0.315)
Treatment*10			-0.811***			-0.264	-0.549**
			(0.260)			(0.252)	(0.263)
Treatment*11			-0.922***			-0.321	-0.609*
			(0.273)			(0.302)	(0.330)
Control variables	No	Yes	Yes	No	Yes	Yes	Yes
Observations	871	871	871	871	871	871	871
No. of schools	146	146	146	146	146	146	146

Includes only schools participating in the 2004-survey.
Robust standard errors (clustered on the school level) in parentheses.
School fixed effects, a constant term and time dummies are included in all regressions.
Weighted regressions using the number of girls/boys/total students as weights.
*** p<0.01, ** p<0.05, * p<0.1