Preparation time and exam scores

How preparation time affects exam scores.

WORKING PAPER

Simon Søbstad Bensnes *

Department of Economics, Norwegian University of Science and Technology

May 18, 2016

Abstract

In the economics literature there is a significant amount of contributions towards estimating the effect of instruction time and school year length. These papers are aimed at identifying the marginal contribution of increasing the time students spend in a instructive environment for the benefit of identifying fruitful policies. This paper adds to the literature by using what is in effect random variation in students' preparation time prior to high-stakes exams. Explicitly, all Norwegian high school students are notified which exams each student will take at the same at a precise date and time. The variation in the exam date thus give students a random amount of preparation time varying between 5 and 25 days in the data. Using this randomization and administrative data, the study finds that 5 extra days of preparation time increases exams scores somewhere between 4.7 and 7.7% of a standard deviation. The effect is fairly similar between students of different backgrounds and abilities. The effect of preparation time also appears to materialize in the longer-run outcomes.

JEL classification: I20 I21

Keywords: High school, test scores, preparation time.

*Correspondence: simon.bensnes@ntnu.no.

1 Introduction

Economists have devoted significant amounts of research into exploring the role of various inputs in the education sector. Often this is done by comparing the results on high-stake tests for students exposed to various amounts of inputs or different environments. Yet, an important factor in determining students' test scores could be the time they have to prepare for these tests. If the time time students have to prepare for tests is a significant factor, it is also of importance to take it into account when assessing other institutional factors. Additionally, identifying the impact of increased preparation time has policy implications: large effects would be a strong argument for homogenous preparation periods between students. In the absence of homogenous preparation periods, high-stakes tests can be more informative of the preparation time rather than students' abilities in the more extreme cases. Further, differences in test results can emerge between students who are otherwise similar when preparation times differ. Thus, the sorting of students into education programs and institutions later on can be a result of differences in preparation times rather than differences in students' abilities.

This paper offers a new contribution in identifying the relatively unexplored role of preparation time. While previous contributions largely have been devoted to identifying the effect of increased time spent in class or in school, this paper is devoted to identifying the direct effect of increased preparation time before high-stakes tests. The identification is achieved by using the unique national exam system in the Norwegian high school system combined with highly detailed administrative data. Specifically, students in the Norwegian high school system are required to take several exams and are notified when exams will be held and which exams they will take on a specific date each year. Because all exams are not held at the same date and because students take multiple exams, it is possible to use random within-student variation in preparation time to assess the effect of increased preparation time. The results suggests that increasing preparation time from 5-9 days to 10-13 days increases test scores by 4.7% of a standard deviation. The effects is relatively large considering that the increase reflects less than 5 days extra time spent in preparation. Because the preparation time is easily adjusted and well within the control locus of administrators, equalizing preparation time for all students is therefore relatively easy compared to adjusting other factors regarding test-taking.

The study presented here is related to the literature on the effect of instruction time, but differs in a distinct way: While previous contributions have focused on the effects of increasing instruction time across a longer time span such as a year, this study focuses on a short and intense preparation period where students are in charge of managing their own time. This study therefore also sheds some light on how students cope with time management prior to large tests. Consequently, this paper also explores the interplay between students' backgrounds and the effect of preparation time.

The paper is organized as follows. Section 2 introduces related literature. The institutional setting is presented in Section 3, while the empirical strategy is presented in Section 4 with the data. Main results are presented in Section 5 with robustness checks in Section 6. Section 7 shows results when the effects are allowed to differ between students according to their characteristics. Longer-run results are presented in Section 8, while Section 9 summarizes and concludes.

2 Related literature

The endogeneity arising from the selection of students into schools or courses constitutes a serious challenge to causal identification of the causal effect of preparation time. In the absence of national standardized regulation of preparation time one can easily imagine that some schools allow students longer preparation time than others. Such schools could also be more focused on students' scores on national tests. If students' and parents are aware of such discrepancies between schools a natural consequence could then be sorting of academically stronger students to these schools. Estimating the effect of preparation time in such an environment would then cause the results to be biased upwards, and consequently be misleading. In the literature on instruction time, a similar sorting problem would occur. To combat the arising endogeneity problem, economists have applied different techniques. Hansen (2011) used data on mandatory school closings due to snowfall in Maryland and Colorado and find that 5 extra days of instruction increases test score between 0.05 and 0.15 standard deviations. Fryer Jr. (2012) finds that implementing a variety of measures in low performing schools in Houston, including increasing instruction time, significantly increased students performance. Specifically he finds that increasing instruction time by 25% increased test scores by 0.08 standard deviations.

Other studies have cross country evidence to estimate the relationship between instruction time and test scores. Wößmann (2003) used TIMSS data including 39 countries, and found that increasing instruction time by 1 standard deviation, or a little less than 6 days, increased test scores in mathematics by 0.025 standard deviations. Lavy (2015) finds that when student fixed effects are employed, increasing instruction time by one hour per week is estimated to increase students test scores by around 0.06 of a standard deviation.

To my knowledge there is only one previous study that directly addresses the effect of preparation time on students' outcomes. In Norwegian lower secondary education students are required to take a mandatory written exit exam at the end of 10th grade. Students' are randomly assigned to take an exam either in languages or in mathematics. Falch, Nyhus, and Strøm (2014) use this information to assess whether there are long term effects on students' enrollment probabilities in tertiary education. Because students have around five days of preparation time, students who are assigned to take the mathematics exam effectively spend five more days studying mathematics than other students. The authors estimate that the causal effect of this preparation period in mathematics relative to languages increases the probability that a student enrolls in tertiary education by around 0.8%-points. This study differs from this previous contribution in some important aspects. First, the estimates presented in Falch et al. (2014) were targeted in identifying the effect of a intense preparation period for a mathematics exam relative to an exam in languages on longer-term outcomes. This study, on the other hand is focused on the effect of preparation time on test scores. Second, This study use data on students in high school. Because high school students are older and because the tests in high schools are generally used as a placement tool in tertiary education around the world, the policy relevance of the findings presented here might be greater. Third, this study employs a student fixed effects framework that was not feasible in Falch et al. (2014). Using student fixed effects effectively nets out any differences between students that could be arbitrarily correlated with preparation time. Thus, this study introduces new evidence, estimated in a new framework on the effects of preparation time on test score results.

3 Institutional background

This section builds on Bensnes (2016). The Norwegian school system consists of ten years of compulsory schooling, starting the year students turn six, and an elective high school education. It is not possible to fail a class in mandatory schooling, implying that grade repetition is practically non-existent and that nearly all students graduate from mandatory schooling at age 16. In this study, all students who did not finish mandatory schooling at the normal age are dropped from the sample. Roughly 95% of students choose to enroll in elective high school education the fall after graduating from mandatory schooling. High school is tracked and consists of 12 tracks that can be grouped into two broad categories: the academic track and the vocational track. Exams in the vocational track often have a practical part and it is not possible to identify which exams include this partitioning. For this reason, only students in the academic track are included in the analysis. The academic track lasts three years and grants graduating students eligibility to apply to higher education. Of the students choosing to enroll in high school the year they turn 16, roughly 50% opt for an academic track. In the academic track, exams are either oral or written and can be identified as such. For comparability to other studies, and because oral examinations are likely to be influenced by non-academic characteristics for the student, only written exams are used in the analysis. The curriculum is comprehensive and standardized at the national level, and written exams are standardized for each school year.

Students in the academic track in the Norwegian high school system are all required to take a written exam in Norwegian language. Besides this exam students are further required to take randomly drawn written exams in their second and third years. In the second year all students are required to take one exam that is either written, oral, or practical. In the third year the number and forms of exams students are required to take varies between sub-specializations in the academic tracks. The majority of students are required to take two randomly drawn written exams and one oral exam, while the remaining students are drawn to take two to three exams that are either written or oral. Finally, around 20% of students are also randomly drawn to take an additional exam in the first year of high school, which can be in any of the three forms.

The exams in the Norwegian high school system are high-stake for three reasons. First, if a student fails an exam they are required to re-take the exam if it is in Norwegian languages. If it is a randomly drawn exam they will be redrawn to take a new exam in either the same or a different course the next semester. Second, in order to graduate students have to pass all exams. Third, students compete for placement in higher education based on the average of their course grades and their exam grades[1]. The fact that tests are high stake suggests that students will utilize their assigned preparation time to the best of their abilities.

4 Data and empirical strategy

4.1 Dependent variable

The main dependent variable in the analysis will be the grade awarded on written exams for students enrolled in the academic track. In this study, individual grade records are collected from register data made available by Statistics Norway, covering the period 2008 to 2011. The data contains identifiers of the course for which the grade is awarded. The exam grades are merged to a list of exam dates from the Ministry of Education. Exam scores range from 1 as the lowest to 6 as the highest and are distributed in a bell shape with 4 as the median grade. Exams are anonymized and are graded for all students in the course-class by two external evaluators who teach the same course at a different high school.

In the data, the number of exams each student takes in high school varies somewhat due to three factors. First, some students in the sample drop out of high school and therefore do not take the number of exams required to graduate. Second, the number of exams taken is not the same for all students who graduate due to exemptions and because a minority of students are randomly assigned to take a written exam the first year in high school.[2] Finally, some students retake exams to achieve a passing grade or a higher score. The student fixed effects framework used in this paper requires students who took at least two exams. Students who took only one exam are consequently dropped from the sample. Students who retake exams in order to improve their grade are marked as such; results from second attempts on any exam are also dropped from the sample. It is important to note that although retake exams are marked in

[1] A typical student has around 20 course grades and 4 exam grades, including oral exams. Thus, each exam counts for around 4% of each student's application score

[2] Students can apply to be exempt from taking the exam in one of the two written forms of Norwegian if they have a native language that is not Norwegian or if they have sufficient difficulty reading.

the data, make-up exams are not and results from make-up exams are thus not distinguishable from results from ordinary exams. In the final dataset, the majority of students took two to five exams, with the median student taking three exams. This is consistent with the normal number of exams students take in high school given student attrition.[3]

4.2 Exam preparation time

The strategy utilizes within-student variation in the exam preparation time. This variation is identified by how exam announcements each year are made. Specifically, after the Ministry of Education has set up a time schedule for when each exam is to be held, each school is sent this schedule along with an announcement time. At the specified announcement time every school is obligated to announce to its students which courses they are to be examined in, and at what time the exam is to be held. Schools can therefore not in any way alter the timing of exams or when students shall be notified. The number of days from the announcement to the exam dates ranges from 5 to 25 days, with a student level average of 13.5 days. In the baseline model the exam preparation time is divided into groups based on the student level distribution of preparation time, but other specifications are explored. There are two main advantages of grouping preparation time. The first is that it increases precision of the point estimates with more observations behind each variable of interest. Second, it does not impose any functional form such as a linear relationship between preparation time and exam scores. A potential drawback of this modeling is that it might drive the estimated effect, a threat that is closer explored in the robustness section.

Because students follow different subjects, all students within a school are generally not tested on the same days. The assignment of students to exams is random given the course composition of each student. That is, given the the courses the student has chosen as elective courses, the expected preparation time will be random. Students who are sick or otherwise unable to show up for the examination in a course they are picked to take, and can provide a medical certificate from a physician stating the reason for their absence, are required to take a

[3]That is: one exam in a randomly picked elective course the second year and written exams in two randomly picked elective courses in addition to one written exam in Norwegian languages in the third year. Elective courses are typically science, mathematics, foreign languages, or social sciences.

make-up exam the following fall semester. The total number of make-up exams constitute less than 1% of all exams (Bensnes, 2016).

4.3 Identification strategy

The identification strategy can be presented as in equation (1). The dependent variable is the exam score in course c for student i taken in the year y. C is a constant, and η_i, γ_c and θ_y are student, course and year fixed effects respectively. X_{iyc} is a vector of student observables that vary between exams. It includes the number of exams the students take the same year and the number of the exam within the year in chronological order. ϵ_{iyc} is a random idiosyncratic error. The coefficient of interest is β, which measures the effect of extra exam preparation time. In the main specification functional form will be given by three dummies for the preparation time being in the second to forth quartile, with the first quartile as the reference category. This definition will be challenged in Section 6 which shows that the results are robust to alternative definitions.

$$\text{Exam score}_{iyc} = C + \beta f(\text{Preparation time}_{iyc}) + \eta_i + \gamma_c + \theta_y + X_{iyc} + \epsilon_{iyc} \tag{1}$$

The main identifying assumption behind the strategy is that preparation time is in effect random given the control variables. Given that both the announcement time and the exam dates are set by the Ministry of Education this seems like a plausible assumption. The assumption is even more likely to hold when estimations includes fixed effects in several dimensions. First, consider a case without course fixed effects. If the exam in a course is systematically placed towards the end of the exam period every year, the preparation time for this course will be longer than for the average course all of the three years in the data. If this course also has a higher return to time spent in self-study, the estimated effect of preparation time will be upward biased. Thus, including course fixed effects removes this potential bias from the estimation. Second, consider a case without student fixed effects. Even when including course fixed effects, some students might randomly receive longer average preparation time periods than other students.

These students might also be better at managing time in this preparation period. If this is the case the estimated effect of increasing preparation time might be biased away from zero and exaggerate the effect. When student fixed effects are included, the estimated effect is only based on random variations within the students. Thus, if some students are better at managing their self-study time than others, the results will not be influenced by this heterogeneity in students.

5 Main results

Table 1 shows the main results when estimating equation (1) with various number of controls. All estimations include student fixed effects. Column (1) shows the most parsimonious estimation when only the number of days of preparation time and student fixed effects are included. In this simple model there is no detectable effect of increasing preparation time. However, the model estimated in Column (1) does not take into account the potential problems with some exams sorting towards the beginning or the end of the preparation period.

Moving on to Column (2) the estimated model includes dummies for the number of exams the student has in the exam period and the number of exams the student has had within the same exam period prior to the exam. The estimated effect is significant and is increasing in the preparation time. Including controls for the number of exams the student has within the same exam period and the order of exams is important as both are likely to influence the time a student spends on studying a specific subject. However, without course fixed effects there is still significant room for bias due to the mechanisms described above.

Column (3) is an expansion of the model estimated in Column (2) and includes year and course fixed effects. The exam year is a relevant control insofar the average thoroughness of exams could vary across years. As can be seen from the table the estimated effect appears to be non-linear. Increasing the preparation time from 5 to 9 days to 10 to 13 days, that is the first to the second quartile in the student level distribution of preparation time, increases the exam score by .0547 grade points or 4.7% of a standard deviation.[4]. Increasing preparation time from the third to fourth quartile adds little if anything to the students' exam grade relative to moving from the first to the second. An effect like this could arise if there is a limit to how

[4]In the sample the average exam grade is 3.252 with a standard deviation of 1.154

much students are able to prepare for their exams.

Last, Column (4) takes the number of controls one step further and includes the teacher assessed course grade as a control. The teacher assessed grade has a very high predictive power, but does not reduce the point estimates. If anything the effect is stronger when the teacher assessed grade is included as a control. The estimated effect of increasing preparation time now shows a strong concave relationship with increasing effect up until 18 days of preparation time and no extra effect when preparation time increases further. It should be noted that it is not entirely clear if the teacher assessed grade is set prior to the date of the exam cases. If the teacher assessed grade is set after the exam is held students who perceive that they performed poorly might influence the teacher's assessment and thereby causing a two-way causality. Because of this potential problem the teacher assessed grade will not be included in specifications in the baseline model.

Summarizing, results in Table 1 shows that it is important to control for not only student fixed effects but also course fixed effects because the data contains only a few years. Within the time period it is therefore not unlikely that exams in some courses are allocated to the end or the beginning of the exam period. A sorting like this could bias estimates unless it is controlled for. The results also show that preparation time appears to have a non-linear effect on exam scores. Section 6 below will challenge the definition of the variable of interest, preparation time. Moving onwards, the model presented in Column (3) will be referred to as the baseline model.

Table 1: Main results

	(1)	(2)	(3)	(4)
10 to 13 days of preparation time	0.0203	0.00433	0.0547***	0.0507***
	(0.0134)	(0.0142)	(0.0150)	(0.0151)
14 to 18 days of preparation time	0.0144	0.0476***	0.0792***	0.0818***
	(0.0108)	(0.0151)	(0.0190)	(0.0189)
19 to 25 days of preparation time	-0.00901	0.0684***	0.0888***	0.0854***
	(0.0142)	(0.0209)	(0.0229)	(0.0222)
Teacher assesed course grade				0.349***
				(0.00610)
Constant	3.244***	3.302***	3.713***	2.236***
	(0.00716)	(0.0101)	(0.0491)	(0.0570)
Observations	277,510	277,510	277,510	277,510
R^2	0.656	0.657	0.678	0.705
Course FE	No	No	Yes	Yes
Year FE	No	No	Yes	Yes
Number of exams same year	No	Yes	Yes	Yes
Number of exams previously same year	No	Yes	Yes	Yes
# Students	100533	100533	100533	100533

Standard errors clustered by high school in parentheses. *** p<0.01, ** p<0.05, *** p<0.1. Source: Statistics Norway

6 Robustenss

The main results presented in the previous section groups preparation time in intervals. Following the results, an immediate question is whether the definition of these variables are driving the results. In particular this is important as the results in Table 1 hints at a non-linear relationship between preparation time and exam scores. This section addresses this concern by presenting results with alternative model specifications. Table 2 mirrors Table 1 with the exception that the preparation time is included as a linear and squared term of the number of days preparation time. As in Table 1 there is no detectable effect in the most parsimonious model. However, when the more controls and fixed effects are added a similar result to the one found above emerges. Strikingly, when moving from the specification in Column (2) to Column (3) the signs on the variables of interest is reversed. This again shows the importance of controlling for course fixed effects. The results in Column (3) shows that increasing the number of days preparation time by 5 increases the exam score by 0.015 grade points or 1.2% of a standard deviation[5]. This effect is somewhat smaller than the estimated effect in Table 1, but shows a strong positive and concave relationship in a quite different model, and is thus supportive of the estimates in Table 1.

[5]The average exam preparation time is 13.6 days, with a standard deviation of 4.9.

Table 2: Robustness

	(1)	(2)	(3)	(4)
Days of preperation time	-0.00257	-0.0251***	0.0181***	0.0198***
	(0.00523)	(0.00560)	(0.00623)	(0.00625)
Days of preperation time, squared	7.66e-05	0.00103***	-0.000468**	-0.000547***
	(0.000186)	(0.000199)	(0.000198)	(0.000202)
Teacher assesed course grade				0.350***
				(0.00610)
Constant	3.271***	3.447***	3.613***	2.127***
	(0.0335)	(0.0374)	(0.0553)	(0.0611)
Observations	277,510	277,510	277,510	277,510
R^2	0.655	0.657	0.678	0.705
Course FE	No	No	Yes	Yes
Year FE	No	No	Yes	Yes
Number of exams same year	No	Yes	Yes	Yes
Number of exams previously same year	No	Yes	Yes	Yes
# Students	100533	100533	100533	100533

Standard errors clustered by high school in parentheses. *** p<0.01, ** p<0.05, *** p<0.1. Source: Statistics Norway

In the Norwegian high school system students have more several exams in their last school year. Thus, when students are informed which dates and courses they are to be examined in, they will also have to divide their preparation time between the exams. A student who has three exams might therefore choose to focus on the exam that comes first in the beginning of the preparation period, and leave work on the remaining exams for the later part of the preparation period. If this is the case, one could argue that the relevant measure of preparation time is the number of days since the previous exam rather than the total number of days between the exam and the announcement date. To meet this potential problem I include both a linear term and dummies for the number of days since the previous exam in Columns (1) and (2) of Table 3. The estimations also include a dummy for whether the exam is the first one in the school year. The point estimates changes somewhat relative to the main results reported in Column (3) in Table 1, but remain statistically indistinguishable. This suggests that the model reported in Table 1 includes the relevant measure of preparation time, and that the students are relatively adept at dividing their preparation time between the exams.

Table 3: Adding days since previous exam

	(1)	(2)
	Linear	Dummies
10 to 13 days of preparation time	0.0547***	0.0571***
	(0.0150)	(0.0150)
14 to 18 days of preparation time	0.0799***	0.0477**
	(0.0194)	(0.0204)
19 to 25 days of preparation time	0.0899***	0.0535**
	(0.0244)	(0.0257)
Constant	3.688***	3.723***
	(0.174)	(0.0488)
Observations	277,510	277,510
R^2	0.678	0.678
Course FE	Yes	Yes
Year FE	Yes	Yes
Number of exams same year	Yes	Yes
Number of exams previously same year	Yes	Yes
# Students	100533	100533

Standard errors clustered by high school in parentheses. *** p<0.01, ** p<0.05, *** p<0.1. Source: Statistics Norway

7 Heterogeneity

7.1 Heterogeneity by student characteristics

Taking the results so far as causal, one might ask if there is an underlying heterogeneity in which students benefit from increased preparation time. The results reported so far show that otherwise similar students will receive different exam scores if their preparation periods differ, even by relatively small amounts. However, the estimations might conceal an underlying heterogeneity as the effect is forced to be the same for all students. The effect of preparation time might therefore differ substantially between students with various backgrounds. Finding such differences would further substantiate the importance of homogenous preparation periods for students. It can also shed light on how students with varying characteristics are able to take advantage of self-study time in a more general perspective.

Table 4 reports results when the data is split in four different ways. The first two columns report estimation results when the sample is split according to the average teacher assessed grade, the first column is for students who have a GPA that is higher than, or equal to, the median, and the second column shows estimates for the rest of the sample. Splitting the sample in this way allows for heterogenous effects across all the independent variables, which is a benefit relative to a model with interaction terms for the variables of interest. The reason the sample is split according to GPA is that one might expect that students who are on average performing better in school will be better able to further increase their relevant skills. One potential mechanism in play here could be that high performing students have some general skills that makes them better at managing their study time. On the other hand, high performing students might have a more limited opportunity to increase their skills: if they are already performing on the upper part of the scale it might be harder to further increase their skills as marginal returns to studying could be decreasing. The results show that the students who are performing in the upper half of all students, on average, have a higher return to extra preparation time: Particularly, when moving from the first to the second quartile of preparation time the students are in the upper part of the GPA distribution have a higher return to preparation time than the rest of the students, and the effect persists throughout to the longest preparation times although it diminishes. Note that the difference in the point estimates is statistically significant only for the first quartile, suggesting that the returns to extra preparation time is decreasing at different rates for different groups of students.

Moving on, Columns (3) and (4) estimates the baseline model for girls and boys separately. Often girls are found to outperform boys in school, this is also the case in the Norwegian school system. Yet, it is not clear ex ante that girls should benefit more from extra preparation time than boys. However, the results suggest that girls might benefit somewhat more than boys when the preparation time increases from the first to the second quartile with little or no difference from further increases in preparation time. This is a similar pattern to what is found in Columns (1) and (2).

The four last columns divide the sample by socio-economic background. The reason for this part of the analysis is that students who come from a more resourceful background might receive more guidance at home in preparing for their exams. A longer exam preparation period might thus allow for even more assistance from parents. Columns (5) and (6) estimates the effect separately for students whose parents have an average income above and below the median in the years 2007-2008[6]. Columns (7) and (8) divide the sample by students for whom at least one parent has a college degree, and students for whom neither parent has any degree beyond a high school diploma. In neither of the sample splits there is any detectable difference for any amount of preparation time. This suggests that exam preparation time has the same positive effect on exam scores for students regardless of their socio economic background.

Summing up, there appears to be some heterogeneity in the effect of preparation time between high and low performing students and between the genders. There is no traceable heterogeneity between students from different socioeconomic backgrounds. Thus, the heterogeneity analysis suggests that increasing preparation time further strengthens the differences in performance between the high and low performing students and between the genders.

[6]These years are chosen because they are the two last years prior to students taking exams

Table 4: Heterogeneity by student characteristics

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	High GPA	Low GPA	Girls	Boys	High parental income	Low parental income	High parental education	Low parental education
10 to 13 days of preparation time	0.0749***	0.0229	0.0675***	0.0372*	0.0519***	0.0580***	0.0502***	0.0614***
	(0.0174)	(0.0187)	(0.0173)	(0.0190)	(0.0187)	(0.0179)	(0.0179)	(0.0183)
14 to 18 days of preparation time	0.0813***	0.0688***	0.0672***	0.0932***	0.0811***	0.0791***	0.0763***	0.0826***
	(0.0240)	(0.0221)	(0.0222)	(0.0245)	(0.0231)	(0.0228)	(0.0225)	(0.0232)
19 to 25 days of preparation time	0.0990***	0.0710***	0.108***	0.0590**	0.0714**	0.108***	0.0801***	0.101***
	(0.0273)	(0.0271)	(0.0275)	(0.0280)	(0.0289)	(0.0266)	(0.0276)	(0.0277)
Constant	4.160***	3.256***	3.752***	3.675***	3.870***	3.561***	3.865***	3.506***
	(0.0639)	(0.0487)	(0.0554)	(0.0701)	(0.0682)	(0.0508)	(0.0602)	(0.0564)
Observations	142,472	135,038	153,461	124,049	138,863	138,647	161,951	115,559
R^2	0.572	0.576	0.676	0.682	0.669	0.679	0.667	0.669
Course FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Number of exams same year	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Number of exams previously same year	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
# Students	52310	48222	55828	44704	49518	51014	58310	42222

High GPA is defined as the student having an average teacher assessed grade above the median. High parental income is defined as parental income in the years 2006-2008 being higher than the median. High parental education is defined as the highest achieved education of either parent being tertiary education. All regression pairs are mutually exclusive. Standard errors clustered by high school in parentheses. *** $p<0.01$, ** $p<0.05$, *** $p<0.1$. Source: Statistics Norway

7.2 Heterogeneity by course type

Although the results so far establishes a robust connection between preparation time and exam scores, the results have not addressed potential differential effects by course type. A general finding in the economics of education literature is that interventions tend to affect mathematics and science results more than results on reading and languages. Consequently, one might expect the effect of preparation time to be dependent on which course students are preparing for. One hypothesis could be that students are able to improve their skills in more quantitative courses with more ease compared to less quantitative courses. To explore this hypothesis I now turn to estimations within students who chose to take courses in three different fields. Each course is marked as in one of three categories: mathematics and natural sciences, languages, social science and others. Then separate regressions are run for students in each group. Running regressions separately has the advantage that it also takes into account potential unobserved factors that varies across time within students, and is correlated with students' choice of courses. It also allows the effect of preparation time differ across students who chose subjects within the varying fields: Students who chose natural science and mathematics courses could be better at preparing for exams for any given preparation time. The results from this approach are reported in Table 5.

Column (1) of Table 5 show results using a model similar to the model estimated in Table 2. The baseline model is not possible to estimate here because it is too demanding for the data. For students who are examined both in courses in science and mathematics and other courses, the effect of preparation time is the same, regardless of which type of course they are examined in: Neither of the two interaction terms in Column (1) are significant. The same pattern emerges for the other fields of study with the exception of social sciences. Yet, the coefficients on preparation time and its square are not significantly different across the models. Thus, it appears that the effect of preparation time is independent of which field of study the exam is. This is somewhat surprising as Falch et al. (2014) finds differential effects between math exams and other exams. However, this study use within student variation in preparation time to assess its effect on exam scores rather than estimating the effect of having an exam in a course per se.

Table 5: Heterogeneity by course type

	(1) Science and math	(2) Languages	(3) Social sciences and other	(4) Norwegian
Days of preperation time	0.0295***	0.0177**	0.00730	0.0224***
	(0.0110)	(0.00853)	(0.0135)	(0.00841)
Days of preperation time, squared	-0.000732**	-0.000604**	0.000114	-0.000572**
	(0.000352)	(0.000258)	(0.000464)	(0.000250)
Days of preparation time x sciences	-0.00670			
	(0.0163)			
Days of preparation time x sciences, squared	0.000359			
	(0.000575)			
Days of preparation time x languages		-0.0108		
		(0.0117)		
Days of preparation time x languages, squared		0.000719*		
		(0.000390)		
Days of preparation time x social sciences and other			-0.0194	
			(0.0182)	
Days of preparation time x social sciences and other, squared			0.000142	
			(0.000591)	
Days of preparation time x Norwegian, squared				-0.00569
				(0.0191)
examprepperiod_norsk_sq				0.000356
				(0.000660)
Constant	3.661***	3.668***	3.553***	3.640***
	(0.0795)	(0.0683)	(0.0847)	(0.0580)
Observations	145,689	264,320	108,299	250,442
R^2	0.667	0.660	0.633	0.645
Course FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
Number of exams same year	Yes	Yes	Yes	Yes
Number of exams previously same year	Yes	Yes	Yes	Yes
# Students	49101	89136	32271	78248

Standard errors clustered by high school in parentheses. *** p<0.01, ** p<0.05, *** p<0.1. Source: Statistics Norway

8 Longer-run

After establishing that preparation time has a significant positive effect on exam scores, this section goes one step further and asks: given the estimated effects, are there any effects on the longer-run effects on students' outcomes? The question is of great relevance because it is not clear if the short-run effects are large enough to materialize in a longer-run impact. If the relatively large differences in preparation time in found in the data are too small to affect longer-run outcomes, one could argue that preparation time has no effect on the relevant outcomes for the student, even if there is an effect on grades. To explore this question I use data on enrollment in tertiary education. The methods here are similar to those applied in Bensnes (2016), which is exposed to the same challenges as this paper using the same data to answer a different question. The core problem in estimating a longer-run effect of increased preparation time is that preparation time varies within the individual, while the enrollment decision is only observed once. Therefore, the student fixed effects approach employed above is not feasible. Instead, two alternative methods are employed. The first one is based on a two-step procedure. In the first step I use the data on enrollment in tertiary education to estimate a descriptive relationship between the average exam grade and enrollment for high school graduates. Because it is likely that there is some sorting of students between applying to tertiary education and not depending on background characteristics and unobservables the estimated relationship does not reflect a causal relationship, but it does give an impression of the connection between exam grades and enrollment. In the second step I combine the estimated effect of increased preparation time from the baseline model in the main results with the descriptive relationship estimated in this section. The combination of these estimates can then be used to predict the longer-run effects of increased preparation time. This approach is not ideal as the relationship between grades and enrollment is merely descriptive, as some graduates who have high enough grades to enroll chose not to apply. Thus, there is a selection issue that could distort the estimated relationship. However, in the absence of data on which students apply, it is not possible to estimate a less biased relationship.

The second approach is a reduced form approach where the enrollment decision is regressed against the average preparation time each student has prior to her exam. In one aspect this approach is less satisfying than the first method as it is to a smaller degree able to take into account the non-linearity found in the main results. On the other hand a benefit is that it allows for a reduced form effect that is not relying on estimating a descriptive relationship between the average exam grade and enrollment. Considering that both methods have strengths and weaknesses results from both methods are reported.

Columns (1) and (2) in Table 6 show estimates for the descriptive relationship between grades and enrollment in tertiary education. The first column shows that, given the average course grade, an increase in the average exam score of 1 standard deviation increases the probability that a student enrolls in tertiary education by about 3.2%-points[7]. Recall that the baseline model predicts that increasing the preparation time for an exam will increase the exam grade by 4.7% of a standard deviation. Consequently, the estimates in Column (3) of Table 1 suggests that if a student goes from having a preparation time for all exams in the first quartile to having all exams with a preparation time in the second quartile, her average exam score should increase by 4.7% of a standard deviation. The effect on enrollment in tertiary education will then be about 0.1%-points, a relatively small effect[8].

Column (2) estimates a similar relationship, but the outcome is enrollment in a science program rather than in any tertiary education program[9]. The regression includes all students who enrolled in tertiary education. This is of interest for two reasons. First, the requirements to enroll into natural science-programs are generally more stringent than other programs. Therefore any effect on the probability of enrolling in any of these programs could reflect an effect on the sorting of students across programs, and not only on whether students enroll in tertiary education or not. Second, increased science and mathematics skills in the population is associated with higher economic growth (Hanushek & Woessmann, 2012). The results suggest that, given the average course grade, a one standard deviation increase in the average exam grade should increase the probability that a student enrolls in a science program rather than any other program in tertiary education by 2.7%-points. This is a large effect given that only around 14% of students enroll in a science program.

The reduced form estimates are reported in Columns (3) and (4). In Column (3) the outcome is whether a students enrolls in tertiary education, while in Column (4) the outcome is whether a student enrolls in a science program rather than another program given enrollment in university.

[7]The student level average exam score is 3.28 with a standard deviation of 0.9829

[8]$0.047*0.9829*0.032\approx0.001$

[9]The Norwegian standard for education classification is applied here. The group includes, mathematics, natural sciences, information technology, and engeeniering

Starting with Column (3), a one standard deviation increase in the average preparation time is estimated to increase the probability that a student enrolls in tertiary education by about 0.8% points[10]. While the estimates in Column (4) suggests that increasing the the average preparation time by one standard deviation increases the probability that a student enrolls in a science program rather than any other program conditional on enrollment is 1.7%-points. The fact that the signs are reversed between Columns (3) and (4) might reflect that the reduced form model has a limited capability to take the relevant non-linearities into account. Thus, the estimates should be interpreted with caution. Yet, the estimated effect of increasing preparation time is sensical and fairly similar to the estimates in the two-step procedure.

Taken together, the two procedures estimated in Table 6 suggests that increased preparation time has longer-run consequences for students by both increasing the probability that students enroll in tertiary education, and by affecting which programs they enroll in. Assessing the implications of the results, it is relevant to contrast the impact of high school exams in Norway to those in other systems. Even though the application score for enrollment in tertiary education is determined by the teacher assessed grades and exam grades, other systems have singular exit exams or put more weight on exams when ranking students applying for tertiary education enrollment. In such systems, differences in preparation periods could play a more significant role in shaping the longer-run outcomes of students.

[10]The student level average preparation time is 13.59 days with a standard deviation of 2.36

Table 6: Longer run

VARIABLES	(1) Enrolling in Tertiary education	(2) Enrolling in STEM in Tertiary education	(3) Enrolling in tertiary education	(4) Enrolling in STEM in Tertiary education
Average preparation time			0.0116**	-0.0271***
			(0.00455)	(0.00592)
Average preparation time, squared			-0.000288*	0.00116***
			(0.000155)	(0.000198)
Average course grades	0.0853***	0.0265***	0.108***	0.0454***
	(0.00346)	(0.00391)	(0.00238)	(0.00266)
Highest parental education: High school	0.0257***	-0.00369	0.0253***	-0.00256
	(0.00810)	(0.00813)	(0.00816)	(0.00806)
Highest parental education: Bachelors or similar	0.0167**	0.0276***	0.0178**	0.0305***
	(0.00773)	(0.00799)	(0.00780)	(0.00788)
Highest parental education: Masters or more	0.00543	0.0850***	0.00786	0.0888***
	(0.00897)	(0.00967)	(0.00906)	(0.00966)
First generation immigrant	0.157***	0.0817***	0.153***	0.0732***
	(0.0115)	(0.0111)	(0.0115)	(0.0111)
Second generation immigrant	0.184***	0.0447***	0.180***	0.0378***
	(0.0104)	(0.0108)	(0.0105)	(0.0110)
Girl	0.0653***	-0.197***	0.0643***	-0.195***
	(0.00404)	(0.00519)	(0.00407)	(0.00513)
Both parents working	0.0153	0.00488	0.0168	0.00779
	(0.0106)	(0.0116)	(0.0106)	(0.0115)
One parent working	0.00647	0.00126	0.00754	0.00335
	(0.0109)	(0.0118)	(0.0109)	(0.0117)
Average parental income 2007-2008	1.54e-08**	1.76e-09	1.55e-08**	1.53e-09
	(6.44e-09)	(8.21e-09)	(6.42e-09)	(8.21e-09)
Average exam score	0.0324***	0.0271***		
	(0.00331)	(0.00441)		
Constant	0.0421**	-0.0244	-0.0388	0.135***
	(0.0187)	(0.0223)	(0.0365)	(0.0443)
Observations	66,570	44,433	66,570	44,433
R^2	0.271	0.124	0.270	0.125
School by Cohort FE	Yes	Yes	Yes	Yes
Number of exams FE	Yes	Yes	Yes	Yes
# Students	1010	982	1010	982

Standard errors clustered by high school in parentheses. *** p<0.01, ** p<0.05, *** p<0.1. Source: Statistics Norway

9 Concluding remarks.

This paper has offered new evidence on the effect of preparation time on students' achievements on high-stakes tests in upper secondary education. Using a unique institutional setting and administrative data, the effects are relatively large and significant, and suggestive longer-run effects. The contribution is unique in the sense that it is the first paper to directly assess the impact of preparation time as opposed to instruction time. While the identification strategy relies on within-student variation and a specific institutional setting, the results are also indicative on how students utilize self-study time prior to important tests. In particular the results are non-linear with differences between students, being more pronounced when the preparation period is relatively short. Considering the that effects are relatively large in Norwegian data, the heterogeniety in the effects could be even more pronounced in countries where students differ more in terms of socio-economic backgrounds. In terms of the longer-run effects established here, effects are likely to be significantly larger in systems where high-stake exams play a larger role in students' later placement in tertiary education.

References

Bensnes, S. (2016). You sneeze, you lose. the impact of pollen exposure on cognitive performance during high-stakes high school exams. *Working paper*.

Falch, T., Nyhus, O. H., & Strøm, B. (2014). Causal effects of mathematics. *Labour Economics*, *31*, 174–187.

Fryer Jr., R. G. (2012). Injecting successful charter school strategies into traditional public schools: Early results from an experiment in houston.

Hansen, B. (2011). School year length and student performance: Quasi-experimental evidence. *Available at SSRN 2269846*.

Hanushek, E. A., & Woessmann, L. (2012). Do better schools lead to more growth? cognitive skills, economic outcomes, and causation. *Journal of Economic Growth*, *17*(4), 267–321.

Lavy, V. (2015). Do differences in schools' instruction time explain international achievement gaps? evidence from developed and developing countries. *The Economic Journal*, *125*(588), F397–F424.

Wößmann, L. (2003). Schooling resources, educational institutions and student performance: the international evidence. *Oxford bulletin of economics and statistics*, *65*(2), 117–170.