Norwegian School of Economics
Bergen, Spring 2021

I don’t believe it:

Correcting students’ misperceptions about peers’ time spent on homework

Julie Olsen & Tobias Lamberg Torjusen

Supervisor: Fanny Landaud

Master thesis, Economics and Business Administration

Major: Economics

NORWEGIAN SCHOOL OF ECONOMICS

Acknowledgements

We would like to extend our deepest gratitude to our supervisor Fanny Landaud. Her advice and guidance has been invaluable in the process of writing this thesis. We are thankful for all comments and suggestions that helped bring this project together.

We would also like to thank Ranveig Falch for supporting us with constructive feedback, proofreading and motivational inputs. Her positive attitude helped us stay motivated throughout the months of working on this thesis.

Finally we would like to thank friends and family for support, proofreading and input to the project.

Norwegian School of Economics

Bergen, June 2021

Julie Olsen   Tobias Lamberg Torjusen

Julie Olsen   Tobias Lamberg Torjusen

Abstract

In this master thesis we investigate the effect of providing students with accurate information about their peers' time spent on homework. We use experimental data collected from two surveys carried out on 10th grade students in Norway. The main survey contained an intervention targeting students who spent below the median time of their class on homework. These students were provided with information about the actual median time spent on homework in their class. The follow-up survey consisted of questions regarding the students' beliefs. A partial population design was utilized in order to capture any spillover effects, in addition to direct treatment effects.

Our main results suggests that the intervention was successful in correcting students' beliefs. Both the reduced form estimation and the instrumental variable estimation suggested a positive treatment effect across our six outcome variables. We used three different specifications, and while we see some differences between them, the main take-away suggests a positive treatment effect.

Our analysis suggests some heterogeneity across students' attitudes, but the evidence is weak. We also check for heterogeneous effects of treatment and spillover conditional on the students' centrality in the peer group. We find some initial differences across these subgroups, but the evidence is ambiguous and does not provide any clear insight into this question.

We recommend further investigation of the direct behavioral changes of such an intervention, as well as more in-depth investigation of the peer effects.

Keywords – Schoolwork, homework, partial population, spillover, instrumental variable, double-lasso selection, information treatment.

Contents

List of Figures

List of Tables

Table 0.1: Commonly used abbreviations

2SLS	Two-Stage Least Square
IV	Instrumental variable
ITT	Intention to Treat
LATE	Local Average Treatment Effect
ATE	Average Treatment Effect
PDS	Post double selection
NHH	Norwegian School of Economics
RCT	Randomized control trial

1 Introduction

Most 10th graders do not enjoy homework. Despite substantial evidence suggesting homework's importance in improving student achievement (Eren and Henderson, 2008), getting a high-schooler to actually put time and effort into the assigned homework can sometimes seem like an impossible task. Researchers have proposed several reasons for this sub-optimal investment, including the opportunity cost of study time (Metcalfe et al., 2019), short-sightedness (Ariely and Wertenbroch, 2002), and underestimation of the expected returns to studying (Ersoy, 2019).

Although these articles raise compelling arguments as to why students underinvest in homework, they fail to thoroughly consider the social side of the students' life. Several articles have studied how individuals are affected by their peers' actions and beliefs (Akerlof, 1991; Falk and Ichino, 2006). 10^{th} graders seem to be particularly susceptible to peer pressure (Brown, 2004). The desire to fit in and conform to the expectations of the friend group affects behavior and choice (Bursztyn and Jensen, 2015). If a student under-reports actual time spent on homework in order to better fit in, it might affect other students' choice as well. This feedback loop could further be enhanced by the friendship paradox (Jackson, 2019), where students with many friendship ties are over-represented in their friends' peer samples. The behavior of the popular students would then be important in the formation of norms regarding homework investment.

A field experiment was conducted by researchers at NHH during the school year of 2020/2021, targeting students who spend less time than the median in their class on homework. An intervention was employed to correct the students' expectations, in order to nudge them to make more efficient choices regarding homework effort.

In this master thesis, we aim to investigate the effect of providing students with correct information about their peers time spent on homework. Our goal is to assess whether the information treatment has any effect on the misperceptions of students. We also want to investigate the role of friendship ties in the spillover effect of this treatment. In order to do so, we first measure the causal effect of the intervention by utilizing an instrumental variable approach. Next, we look for spillover effects, exploiting the partial population design of the experiment. Finally, we measure different network effects, including diffusion and centrality, by using the subjects self-reported friendship ties. We test our results for heterogeneity and robustness by running our analyses on different sub-populations and controlling for potential mechanical issues.

2 Background

2.1 Benefits of Homework

For the purpose of this paper, homework can be defined as any task assigned by schoolteachers intended for students to carry out during non-school hours (Cooper, 1989). Most students, parents, teachers and researchers believe that homework can be an important supplement to in-school academic activities, and that homework is a necessary and valuable part of a student's learning process. Researchers have suggested a long list of positive consequences of homework, both in the academic and non-academic spheres of life (Cooper et al., 2006). Homework generally requires students to complete tasks with less supervision and under less severe time constraints than during school, which is said to promote greater self-direction and self-discipline, better time organization, more inquisitiveness, and more independent problem solving (Corno, 1994; Zimmerman et al., 1996).

Even though our main focus does not involve the effects of homework on achievement, it is still of great importance to our paper. Research on the relationship between homework and academic achievement suggest that doing more homework can have a positive effect on the students' grades (Cooper et al., 2006). This forms the basis of our thesis, as it is imperative that doing more homework yields a positive outcome on achievement and in the development of non-cognitive skills when trying to influence students to do more homework. We consider students who invest relatively small amounts of time to homework, as it might be unclear whether students already spending a large amount of time on homework will benefit from being pushed to do even more.

2.2 Student Effort

A student's study effort is argued to be one of the most important determinants of their human capital (Costrell, 1994), and it is a critical component of their education production function (Stinebrickner and Stinebrickner, 2004, 2008). Studies have demonstrated that students study more when incentives to do so increase (Hirshleifer et al., 2015; Azmat and Iriberri, 2016), and that their beliefs about how much they need to study are often strong predictors of their actual decisions (Stinebrickner and Stinebrickner, 2008). However, previous work has also shown that students often have incorrect beliefs about their own education production function, specifically about returns to their effort (Fryer Jr, 2016; Ersoy, 2019). Because of the importance of student effort and the incorrect beliefs associated with it, understanding how students make their study effort decisions is of high importance for both scholars and policymakers (Rury and Carrell, 2020).

The study effort decisions are also important for the students, as studying more implies less time for non-studying activities such as leisure and work (Stinebrickner and Stinebrickner, 2003; Metcalfe et al., 2019). Thus, students must know their returns to study effort in order for them to make efficient trade-offs between studying and non-studying activities (Rury and Carrell, 2020).

2.3 Adolescence and Peer Effects

In the field of economics, researchers have accumulated large amounts of evidence on the importance of peer effects. Group structures are ubiquitous in education and group composition may have important effects on education outcomes. Furthermore, students find themselves in different classrooms, living environments, schools, and social groups, and are thus exposed to different peer groups, receive different education inputs, and face different institutional environments (Garlick, 2013). Because of this, understanding how social concerns or peer pressure impacts student's beliefs and actions is of high importance for both scholars and policymakers.

Researchers have found that adolescents in a particular peer group exhibit many similarities compared with adolescents in other groups (Nurmi, 2004). This form of homogeneity among individuals in peer groups has been reported in several characteristics, such as aspirations (Kandel, 1978), problem behavior (Urberg et al., 1997) and schoolwork (Cohen, 1977).

The most widely repeated assertions about peer relations during adolescence are that they become increasingly important and occupy an increasing amount of an individual's time (Brown, 2004). Starting from an early stage, children spend an increasing amount of time with their peers both at school and after school (Larson and Richards, 1991), and peer influence arises partly because popular youth often have the power to set styles and determine what activities will be undertaken (Brown, 2004).

3 Literature

3.1 Effects of Receiving Information

Researchers have found that students' beliefs about how much effort they need to put into their schoolwork become more accurate upon receiving information. According to Ersoy (2019), both information about an average individual and anecdotal information moves student's beliefs towards the information provided. Furthermore, students change their study effort in the same direction as the shifts in their beliefs. Further backing this theory, Azmat and Iriberri (2016) argues that information on how students compare to their classmates is relevant when determining how much effort to exert.

Previous research show that providing students with feedback on their relative performance has an impact on their future performance (Azmat and Iriberri, 2010; Bandiera et al., 2015), while another part of the existing literature argues that students exert effort primarily because they are compelled by cultural norms rather than objective rewards (Figlio et al., 2019; Gneezy et al., 2019). Our paper differs in the fact that our main outcome variable is planned time spent on homework, and that our main focus is correcting the student's misperceptions about peers' study effort. The aforementioned papers on the other hand, focuses on returns to study effort and the student's perceived returns. However, our paper will contribute to both parts of the existing literature, as we investigate both direct treatment effects and spillover effects.

Azmat and Iriberri (2010) suggest two alternative explanations for why students would react to the relative performance information. The first being that students might react to additional information because individuals have inherently competitive preferences, or that the presence of relative performance information instigate this type of competitive preferences. In the presence of such competitive preferences, information that allows for social comparison gives students utility from being ahead, and disutility from being behind others.

The second explanation is that individuals' imperfect knowledge of their own ability might lead students to react differently to additional information, such that the information is informative of the student's own ability. An example of this is provided by Rury and

Carrell (2020). In their paper, they study the effect of providing students with information on returns to study effort and find that students who expect to receive low grades may have inflated beliefs about how much effort they need to exert in order to improve their performance. In turn, this leads them to provide effort that is potentially lower than they would if they knew the true returns to effort. That is, if performance is a function of both ability and effort, the self-perceived ability will affect the optimal choice of effort.

Based on these two explanations, all students would either choose high effort when information is provided, leading to an observation of higher performance, or top performing students would choose higher and bottom performing students lower effort, because this information encourages high ability and discourages low ability students (Azmat and Iriberri, 2010).

We see through prior research that the provision of information involving performance feedback allows for social comparison, i.e., individuals can evaluate their own performance by comparing themselves to others (Azmat and Iriberri, 2010). However, social comparison does not only originate from received information through the treatment, but it is also closely connected to the sharing of information among students, and the accompanying peer effects. These peer effects are of great importance to our paper, as they could possibly impact the aforementioned effects of receiving information.

3.2 Peer Effects on Student Effort

Whether or not students would benefit from interactions with other students is an important question in existing research as well as in our thesis. The effect of peers on a student's performance is expressed in the findings of numerous researchers. Carrell and Hoekstra (2010) and Figlio et al. (2019) found that the presence of disruptive peers within classrooms would increase a student's propensity to misbehave and disengage during regular class time. While other researchers found that less disruptive behavior and a sense of futility mediated peer effects on students' academic performance (Avvisati et al., 2013). A third finding is that hardworking peers might serve as role models that inspire other students to put more effort into studying (Hoxby and Weingarth, 2005), and furthermore Pop-Eleches and Urquiola (2011) found that attending a secondary school with high-ability peers increased students' frequency of doing homework after school.

However, these estimates may reflect, but not reveal, behavioral responses that amplify or reduce the impact of educational quality. For instance, these responses might change over time, and thus potentially influence results differently depending on when outcome data are collected (Pop-Eleches and Urquiola, 2011).

Educational researchers have also studied whether the effect of peer composition on achievement is different for students with different academic abilities. Students at the bottom of the test score distribution benefit significantly from the addition of students who are at the 15th percentile of past test scores. Students at the top decile, benefit strongly from the addition of classmates who are also at the top, while achievement for students at the middle tends to be less affected by peer composition (Burke and Sass, 2008; Imberman et al., 2012)

3.3 Friendship Paradox

The friendship paradox refers to the fact that, on average, people have strictly fewer friends than their friends have (Jackson, 2019). In his paper, Jackson suggests two reasons why we should expect more connected individuals to behave systematically different from less connected agents. The first is that people who have the most connections are also the most exposed to interactions with others. This in turn leads to them being most heavily influenced. The second is that if people differ in their taste for different activities, the people who benefit the most from a given activity choose to have the most connections. These two combined lead people's most popular friends to engage the most in a behavior and to bias the overall behavior in the society. Many forms of behavior are peer influenced and driven by people's perceptions of what is normal or acceptable behavior. The impact of the friendship paradox on such behaviors can be seen in a series of studies that students tend to overestimate the frequency of which their peers smoke or consume alcohol and drugs, often by substantial margins (Jackson, 2019). In order for the friendship paradox to have an effect in our case, the more popular students have to be more likely to do more homework influencing their peers to do the same. Alternatively, they have to be more likely to do less homework than the average. If this is the case, the students will be treated, which enables them to spread the information among their connections.

Our study relates to the empirical literature on the diffusion of treatment effects through social networks. In a study on how participation in a microfinance program diffuses through social networks, Banerjee et al. (2013) found that participants were significantly more likely to pass information on to friends and acquaintances than informed non-participants, but also that an individual's decision is not significantly affected by the participation of her acquaintances. The researchers found that the eigenvector centralities of initially informed individuals are significant determinants of the eventual participation rate in a village, while other variations in social network characteristics across villages are relatively insignificant determinants of diffusion. Specifically, they found that individuals who have more participating friends are more likely to participate because they are more likely to hear about it or because they are influenced by the numbers of their friends who participate.

4 Experimental Protocol

4.1 The Norwegian School System

The Norwegian school system is divided into three levels, Primary, lower-secondary and upper-secondary schools and higher education. In Norway, the first 10 years of school are mandatory for all children aged 6-16. These 10 years consists of primary school (1-7$^{\text{th}}$ grade) and lower secondary school (8-10$^{\text{th}}$ grade). All adolescents aged 16-19 also have the right to attend upper-secondary school, but it is not mandatory (Thune et al., 2019). A standard school day for 10$^{\text{th}}$ graders in Norway lasts from approximately 08:30 to 14:00, with small variations between different schools. Homework is assigned by the teachers, and is completed outside of school hours.

4.2 Experimental design

The experiment was conducted by researchers at NHH from the fall of 2020 throughout the spring of 2021. 17 schools were recruited to participate in the study from all over Norway. The experimental program consisted of a main survey, wherein treatment was delivered, and a follow-up survey.

The main survey was distributed to students during school hours, and completed under the supervision of their respective teachers. It included questions regarding the subjects' time use on homework and other activities outside of school hours, as well as questions regarding both the personal and social value of these activities (see Appendix A1.1 for the full survey). The students were also asked to name the other participants from their class, whom they considered to be friends with, and what their belief was regarding their classmates' time spent on homework.

Treatment was delivered to students who during the first questions of the survey reported that they spent less time than the median for their class on homework. Towards the end of the survey, these students were shown a message as seen in Fig. 4.1, informing them that they were among the students who spent the least time on homework in their class. The message also included information about what the median time spent on homework was in their particular class. Immediately after this message all the participating students were asked to report how much time they planned to spend on schoolwork outside of school hours until they finished 10th grade (see Tab. 4.2).

In your class, the majority of the students ([proportion inserted]) spend at least [**median time inserted**] on schoolwork outside of school hours on typical schooldays.

This means that, in everyday life, most students in your class **spend more time than you** on schoolwork.

Figure 4.1: Message displayed to treatment group.

Think about a typical schoolday in the coming time until you complete the 10th grade. If you exclude the time you are at school, how much time per day do you plan to spend on schoolwork?

Alternatives: No time/1-14min./15-29min./30-44min.-/45-59min./1:00-1:14/1:15-1:29/1:30-1:44/1:45-1:59/2:00-2:14/2:15-2:29/2:30-2:44/2:45-2:49/3 hours or more.

Figure 4.2: Question regarding students' planned time to homework.

4.3 Randomization

The randomization was conducted at a class level, defining treatment and control classes based on predetermined stratas. Treatment was then delivered only to students in treated classes who reported below median time spent on homework in their class. This "partial population" design (Avvisati et al., 2013; Moffitt, 2000) makes it possible to capture not only the direct effect of the intervention, but also the spillover effects. The difference in outcomes between below-median students in treatment and control classes captures the effect of being made eligible for the intervention, while the difference between the above-median students in treatment and control classes captures the spillover effects of the intervention.

4.4 Outcome variables

Throughout this thesis we will rely on six main outcome variables. These are constructed from survey data, and measure different aspects of the students' beliefs and behavior.

4.4.1 Main survey

We use two outcome measures from the main survey. These are utilized to detect any direct treatment effects. The first is *Time planned on homework*, and follows directly from the survey. Here the students were asked to report how much time they were planning to spend on homework every day until the end of 10^{th} grade. We also use *Planned homework above median* which tells us whether the students' planned time is above the actual median in their class.

4.4.2 Follow-up survey

In the follow-up survey we are more interested in examining the students' beliefs, rather than their behavior. Our main outcome measure is *wedge* which is a measure for the difference between the students' guessed proportion below median and the actual proportion below median in the main survey. Next, we examine if the probability of students to correctly guess the class median is affected by treatment through the variable *Correct guess*. The two final outcome measures relate to the precision of this guess. The first is *Distance from correct guess*, and measures the difference between the guessed median and the actual median. The final variable, *Absolute distance from correct guess* gives the absolute value for the difference between the guessed median and the correct median. Together, these latter variables should allow us to assess the direction of any over- or underestimation.

4.5 Balance testing

Randomized control trials (RCTs) build upon the assumption that true random assignment of treatment stochastically distributes all baseline characteristics (Mutz et al., 2019). While this does not guarantee perfect distribution of such characteristics, it does allows researchers to make precise quantifiable inferences. What makes random assignment superior to other approaches to inference about causation is the fact that there is an underlying mathematical model supporting the probability of unequal distribution of baseline characteristics. This implies that the researcher is enabled to evaluate the exact probability of imbalances in covariates between treatment and control groups to appear.

To concretize this notion, balance tests of baseline characteristics are usually carried out when reporting on RCTs. The implications of such tests, however, are not entirely straightforward. True randomization eliminates any external influence on treatment indicators, implying that any differences between groups are due to chance. The test statistics from balance testing has the interpretation of the probability that the difference between two groups have occurred by chance, when there in fact is no difference. As noted by Altman (1985) performing such tests "is to assess the probability of something having occurred by chance when we know that it did occur by chance".

In this thesis we take a more pragmatic approach to balancing. We present a table of baseline characteristics with means and differences between treatment and control group, an approach similar to the one advocated by the CONSORT guidelines (Schulz et al., 2010) and APSA standards (Gerber et al., 2014). In addition we present the test statistic of an omnibus test on joint significance, following Hansen and Bowers (2008). This table will serve as a starting point for our discussion. In the formal analysis, we will include only those covariates that, *ex ante*, were argued to be influencing the outcome. This follows the reasoning from Roberts and Torgerson (1999) and Mutz et al. (2019). More detailed discussion of the relevant baseline characteristics follows in the next section, as well as details regarding the selection procedure in chapter 5.

4.5.1 Baseline tests

Table 4.1 shows the result of standard tests for differences in means between treatment and control group. *Panel A* shows the differences for students who spend less time on homework than the median in their class. Only one of the baseline characteristics is statistically different from zero. A coefficient of 0.13 for the row variable *Female* implies that there are 13 percentage points more female students in the treatment group relative to the control group. The estimate is significant at a 1 % level. In addition, *expectations parents > 3* is significantly different across the two groups at a 10 % level. All other baseline characteristics seem to be fairly balanced.

Table 4.1: Differences in baseline characteristics across treatment and control group

	Mean C	Mean T	T-C	(se)	n.obs
Panel A: Below median					
Female	0.29	0.40	0.13***	(0.040)	286
Time homework	13.80	14.60	1.74	(1.537)	287
Dislikes homework	0.86	0.81	−0.01	(0.035)	287
Grades					
Standardized grade math	−0.08	−0.154	−0.08	(0.107)	287
Standardized grade Norwegian	−0.15	−0.29	−0.15	(0.127)	287
Returns to studying Norwegian	0.52	0.50	−0.08	(0.062)	287
Returns to studying math	0.65	0.62	−0.05	(0.070)	287
Attitudes					
Importance grades > 3	0.87	0.80	−0.06	(0.042)	287
Pleasing parents > 3	0.61	0.60	0.04	(0.054)	287
Expectation of parents > 3	0.92	0.85	−0.08*	(0.043)	287
Pleasing teachers > 3	0.48	0.49	0.02	(0.060)	287
Expectations of teachers > 3	0.84	0.83	−0.02	(0.039)	287
Importance of popularity > 3	0.76	0.74	0.01	(0.048)	287
Popularity schoolwork < 3	0.24	0.18	−0.05	(0.043)	287
Friendship					
Number of in friends	3.09	2.96	−0.35	(0.264)	283
Number of out friends	2.92	2.89	−0.09	(0.263)	283
Number of reciprocal friends	6.01	5.84	−0.45	(0.511)	283
Eigencentrality	0.45	0.51	0.03	(0.037)	283
Many friends outside class	0.50	0.36	−0.09	(0.057)	287
Beliefs					
Guessed median - class median main survey	−16.13	−16.88	−2.22	(1.528)	287
abs(guessed median - class median) main survey	19.07	19.77	1.21	(1.069)	287
Wedge main survey	0.31	0.30	−0.01	(0.028)	287
Guessed median = class median main survey	0.11	0.12	0.01	(0.038)	287
Panel B: Above median					
Female	0.61	0.55	−0.08**	(0.033)	504
Time homework	47.27	47.09	−0.26	(1.644)	504
Dislikes homework	0.62	0.61	−0.01	(0.041)	504
Grades					
Standardized grade math	0.07	0.07	−0.04	(0.068)	504
Standardized grade Norwegian	0.19	0.06	−0.20***	(0.066)	504
Returns to studying Norwegian	0.34	0.38	0.06	(0.039)	504
Returns to studying math	0.56	0.59	0.05	(0.041)	504
Attitudes					
Importance grades > 3	0.93	0.86	−0.07***	(0.025)	504
Pleasing parents > 3	0.72	0.76	0.05	(0.037)	504
Expectation of parents > 3	0.92	0.91	−0.01	(0.018)	504
Pleasing teachers > 3	0.60	0.65	0.05	(0.037)	504
Expectations of teachers > 3	0.88	0.88	−0.02	(0.021)	504
Importance of popularity > 3	0.77	0.79	0.03	(0.026)	504
Popularity schoolwork < 3	0.16	0.20	0.04	(0.023)	504
Friendship					
Number of in friends	2.97	2.87	−0.19	(0.196)	494
Number of out friends	3.08	2.91	−0.27	(0.224)	494
Number of reciprocal friends	6.05	5.79	−0.46	(0.413)	494
Eigencentrality	0.47	0.51	0.04	(0.025)	494
Many friends outside class	0.42	0.37	−0.03	(0.042)	504
Beliefs					
Guessed median - class median main survey	−4.25	−4.44	−0.51	(1.898)	504
abs(guessed median - class median) main survey	12.78	13.62	1.23	(1.028)	504
Wedge main survey	0.15	0.12	−0.05*	(0.024)	504
Guessed median = class median main survey	0.28	0.27	0.00	(0.031)	504

Notes: Female is a manually coded variable based on the name of the student. The first and second columns show the mean value of the row variables for the control and treatment group, respectively. The third column shows the estimated coefficient from a regression of the baseline characteristic on treatment status, controlling for strata fixed effects. The fourth column includes robust standard errors, clustered at the class level (corresponding with randomization level). Each row includes a separate regression. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

In *Panel B*, we see a more stark difference between treatment and control group. In total, 4 of the 19 baseline characteristics tested are statistically different between the two groups. Of these, two are significant at a 1 % level, one at a 5 % level and the final one at a 10 % level. While these differences might seem dramatic, we argue that with proper care, we are able to discuss our findings as reliable. In order for us to do so, we rely on two main notions.

First, is the notion that randomized selection does not guarantee equal distribution, only *stochastic* distribution of baseline characteristics (Mutz et al., 2019). This implies that statistical testing of several baseline characteristics are prone to Type I error; the test returns a statistically significant effect when there in fact is no such effect. The probability that at least one of the tests return a significant result increases with the number of baseline characteristics to be tested. We are confident that the randomization was successful; there were no mechanical or technical issues with the implementation of the randomization procedure. As long as the former statement is true, the differences between treatment and control group as shown by the balance test only convey the random nature of stochastic distribution.

In total we have data on 61 different classes distributed across the treatment (29) and control group (32). While increasing this number would increase the probability of a balanced distribution of baseline characteristics, we argue that we still have a decently sized data set to work with.

Second, when conducting our analysis we are concerned about both the efficiency and credibility of our model. With credibility, we refer to the degree in which our results truly reflect the effect of a change in treatment status for an individual. While a randomized and perfectly balanced data set provide strong arguments for credible results, we would nevertheless be interested in arguing for the precision of our findings. The precision may be considered as the propensity for Type II error; finding non-significant results when there in fact are significant effects. A proven way to improve upon a statistical analysis is to include covariates in the analysis. An important prerequisite for improving the model is that the researcher provides some theoretical or empirical evidence, or at the very least well-substantiated suspicions, that the covariate affects the outcome variable of interest. Any covariates which meet these requirements should be included in the analysis in order to increase the precision of the estimates.

The important implication from this discussion is that any such covariates should be included in the analysis regardless of their significance in a balance test (Mutz et al., 2019). In our analysis we include covariates we suspect could affect the outcome of interest, based on *ex ante* discussion. These covariates include indicators for *gender, baseline time spent on homework, attitudes towards homework, grades, perceived returns to studying and popularity measures.* Coincidentally, several of these covariates coincide with the unbalanced baseline characteristics in Table 4.1.

In addition to the individual regressions, we also conducted an omnibus test, following Hansen and Bowers (2008). This method effectively tests for joint significance for all baseline characteristics. Running the test separately for Panel A and Panel B gives a *p-value* of 0.029 and $4.195e-05$, respectively. These low $p-values$ imply that the samples are indeed unbalanced, and confirms our suspicions from the individual balance tests.

In summary, we argue that while our sample seems to be subject to some imbalances in baseline characteristics, careful attention to the issue of balance and randomization combined with meticulous treatment of the covariates should allows us to conduct our analysis with confidence that the estimates it provides are both precise and efficient. Still, we need to carefully consider our choice of estimation model in order for our analysis to be credible.

5 Empirical Approach

The main goals of our empirical analysis are to estimate the causal effect of the intervention, identify any spillover effects from the treated to the non-treated, and study the spillover effects in relation to network characteristics. More specifically, our main outcome variable for the first part of the analysis is *planned time spent on homework*, which captures the students study intentions following the treatment. In the second part of the analysis we will focus on the beliefs of the students regarding their peers' time spent on homework, as reported in the follow-up survey.

5.1 Instrumental Variable approach

When working with a field experiment, as described in chapter 4, we have to consider the possibility of treatment dilution and imperfect take-up (Angrist and Pischke, 2014). In order to circumvent this issue we utilize an Instrumental Variable (IV) approach. This method employs an instrument, in place of the suspected endogenous variable, which allows for only the exogenous part of the explanatory variable to be captured in the model. This method allows us to estimate the local average treatment effect (LATE), as opposed to the intention to treat (ITT) from the reduced form.

Successful IV-estimation requires the use of a valid instrument. There are three main assumptions that has to be satisfied for this approach to be meaningful (Angrist, 2006). The *relevance assumption* requires the instrument to have a significant effect on the instrumented variable. This assumption is trivial in its identification, and can be examined through a regression of the instrument on the instrumented variable. Formally it translates to $Cov(X, Z) \neq 0$

In addition to the relevance condition, the instrument also has to be uncorrelated with the unobserved random effects captured in the regression model, often formulated as $Cov(Z, u) = 0$. This requirement is separated into two assumptions, the *exclusion restriction* and *independence assumption* (Angrist et al., 1996). These assumptions relates to the channels in which the instrument is affecting the outcome variable, and the distribution of the instrument. The *exclusion restriction* requires the instrument to only affect the outcome through the instrumented variable. The *independence assumption* requires random distribution of the instrument.

5.1.1 Instrumental variables in randomized control trial

Randomized control trials are by many considered the gold standard in estimation of causal effects (Angrist et al., 1996), however, this approach requires the researcher to be able to imagine the outcome in the counterfactual situation of no treatment. With random distribution of treatment to a sufficiently large sample, it is possible to argue that the average difference between treated and non-treated corresponds to the average causal effect (Angrist and Pischke, 2014). These types of analyses are usually difficult to perform properly in social sciences, as they require no treatment dilution nor any non-compliance.

Often an alternative approach is to combine RCTs with IV-estimation (Angrist, 2006; Angrist et al., 1996). The novel idea here is to use assigned treatment as an instrument for actual delivered treatment. This allows us to interpret the estimated effect as LATE, which corresponds to the effect of the instrumented variable on the compliers, rather than as an ITT effect.

In this application of the IV framework, the necessary assumptions for a valid instrument becomes somewhat trivial. By design, our instrument is both strongly correlated with the instrumented variable and randomly assigned. This implies that the *relevance assumption* and *independence assumption* are satisfied. We provide evidence for a significant first stage regression in the formal empirical estimation in Appendix A2.1.

The *exclusion restriction* may require some discussion and external motivation in other applications, but for our specific case it is quite trivial. Being randomly assigned to treatment by external researchers should have no impact on the outcome variable of the student whatsoever, if not for delivered treatment. This implies that by design, our instrument is strong and arguably valid for our purposes.

5.2 Empirical model

We specify the identifying model which we estimate separately on students above and below the median of time spent on homework. This approach is similar to the one employed by Avvisati et al. (2013), utilizing the partial population design of the experiment, and allows us to identify the direct effects of being treated by the intervention, as well as the spillover effects of having treated students in your class. The first-stage regression shows the effect of the instrument on the instrumented variable.

$$D_{ic} = \alpha_1 + \phi Z_{ic} + \gamma X_{ic} + \beta_{1c} + e_{1ic} \tag{5.1}$$

Following the naming convention used in Angrist & Pischke (2015), D_{ic} is the instrumented variable, delivered treatment for individual i in class c. Z_{ic} denotes the instrument, and corresponds to the randomly assigned treatment for individual i in class c. X_{ic} is a vector of control variables for individual i in class c, and β_{1c} are dummy indicators for strata fixed effect. e_{1ic} represents unobserved individual random effects.

We get the corresponding reduced form equation by directly regressing the instrument on the outcome variable:

$$Y_{ic} = \alpha_0 + \rho Z_{ic} + \delta_0 X_{ic} + \beta_{0c} + e_{0ic} \tag{5.2}$$

where Y_{ic} represents the outcome measure for student i. X_{ic} is a vector of control variables for individual i in c, and β_{0c} are dummies for strata fixed effects. e_{0ic} is the unobserved individual random effects. The parameter ρ represents the ITT and reflects the effect of being made eligible for treatment.

The fitted values from estimation 5.1 are then substituted into the second stage regression in place of the instrumented variable. This gives the following formal estimation:

$$Y_{ic} = \alpha_2 + \lambda \hat{D}_{ic} + \delta_2 X_{ic} + \beta_{2c} + e_{2ic} \tag{5.3}$$

where Y_{ic} is still the outcome variable, $\hat{D}_{ic}$ is the fitted values from the first stage regression, X_{ic} is the same vector of controls as in the first stage, β_{2c} represents strata dummies, and e_{2ic} is the unobserved individual random effects. The parameter of interest here is λ, which captures the instrumented variable's effect on the outcome variable through the instrument. The estimated λ for students with below median time spent on homework corresponds to the treatment effect, while for students above median time spent, it corresponds to the spillover effects.

We specify three different models. Model (1) is the simplest and only controls for strata fixed effects. Model (2) includes covariates that *ex ante* were argued to potentially affect the outcome measures. Model (3) choose the individual control variables X_{ic} based on a double lasso selection procedure (Belloni et al., 2014). This method effectively allows us to choose the appropriate control variables in a high-dimensional data-set by utilizing a two stage process (Urminsky et al., 2016). First, we identify the covariates that predict the dependent variable, then the ones that predict the independent variable. The final regression model is fitted with the variables that have been estimated to have non-zero effects in the two prior steps.

To avoid issues with incorrect standard errors, we use the built in 2SLS function in STATA to conduct our estimations. In addition, we cluster the standard errors at the class level (Angrist and Pischke, 2008). Randomization of treatment was conducted at class level, and thus the clustering should follow the level of randomization.

5.3 Heterogeneity analysis

We test for heterogeneous results across subgroups by modifying the estimated equations above. We estimate the difference by introducing an interaction term to our estimation. By interacting the indicator for treatment and the characteristic we want to test for heterogeneity across, we get a model that singles out the effect of being treated in a specific subgroup. Below, only the second stage is shown, however in practice we follow the same estimation strategy as above.

$$Y_{ic} = \alpha_2 + \lambda \hat{D}_{ic} + \gamma_2 A_{ic} + \zeta_2 (A \times Z)_{ic} + \delta_2 X_{ic} + \beta_{2c} + e_{2ic} \tag{5.4}$$

Where Y_{ic} is the outcome measure, X_{ic} is a vector of control variables for individual i in class c, β_{2c} controls for strata fixed effects, and e_{2ic} is the unobserved individual random effects. A_{ic} here represents the subgroup identifier. This is a dummy variable that takes the value of 1 if individual i in class c belongs to a specific subgroup, and 0 otherwise. The interaction term is a dummy representing whether the individual belongs to a certain subgroup *and* is treated. The coefficient of interest in this model is ζ_2 which represents the relative difference between treatment effect for individuals in the subgroup and not in the subgroup.

5.3.1 Network effects

In addition to the direct and indirect effects, we are interested in identifying how different network characteristics affect the diffusion of treatment effect. In particular, we want to examine how network characteristics such as eigenvector centrality and degree distribution affect the magnitude of spillover effects. The main idea is that if the treated students in one class are more central in their networks relative to other classes, the spillover effect should be stronger due to a higher degree of interaction between treated and non-treated students.

We test for this by further specifying our heterogeneity analysis. First, we calculate the degree distribution and eigenvector centrality for each friend network, using GEPHI. These variables are then used to construct measures for the centrality of the treated individuals in different classes. Consolidating this with the above approach to heterogeneity analysis allows us to estimate the difference in treatment effect conditional on the initial injection point. This approach bears similarities with Banerjee et al. (2013).

6 Analysis

6.1 Reduced Form Estimation

We start out our analysis by testing for ITT effects by estimating equation 5.2. The dependent variable is different measures for the students' beliefs regarding their future time use on schoolwork. Table 6.1 presents the beta coefficient from the estimations.

All three models (1, 2, 3) suggests that being below median in a class that is made eligible for treatment is associated with an increase in the amount of time planned for homework. The first row represents the mean time planned, while the second row represents the probability for the student to plan more homework than the actual median in the class. There are some differences between the three models. In the first row, we find some difference in the magnitude of the estimated coefficient between model (1) and the other two. In the second row, model (2) reports a slightly higher estimated coefficient relative to the other two. However, the differences are not dramatic, and the significance and general interpretation of the results remain stable across all models.

Table 6.1: Direct intention to treat effects from the main survey

	(1) No controls	(2) Pre-determined controls	(3) Double Lasso Selection
Panel A: Direct effect			
Time planned homework	8.149***	8.457***	8.480***
	(2.295)	(2.811)	(1.921)
Planned homework above median	0.154***	0.240***	0.153***
	(0.058)	(0.060)	(0.056)
N	287	286	286
Panel B: Spillover effect			
Time planned homework	−2.006	−3.208	−4.405*
	(2.988)	(2.640)	(2.264)
Planned homework above median	−0.010	−0.026	−0.031
	(0.029)	(0.029)	(0.026)
N	504	504	504

Notes: Model (1) shows the coefficient from the reduced form regression using only indicators for strata fixed effects as covariates. Model (2) shows the coefficient from a similar regression, but also including pre-determined controls. Model (3) shows the coefficient from a PDS-lasso regression, utilizing a post double lasso selection procedure in order to determine which covariates to include. Robust standard errors clustered at class level in parenthesis below each coefficient. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Considering model 2 (3), the magnitude of the coefficients in the first row implies that being made eligible for treatment is associated with a 8.457 (8.480) minutes increase in average time planned for homework. Further, the second row implies a 24.0 (15.3) percentage points increase in probability for the students to plan more time to homework than the median in their class. Not only do the students plan more time for homework after the treatment, but on average the probability that they plan more than the median in their class is increased. The link between treatment and effect seems to be consolidated through these estimations.

Next, we consider the reduced form estimation of the relationship between treatment and the students' ability to correctly guess their peers' time spent on homework as reported in the follow-up survey. Table 6.2 shows the beta coefficient from estimating equation 5.2 with measures for the precision in students' guesses as the dependent variable.

Table 6.2: Direct and indirect intention to treat effects from the follow-up survey

	(1) No controls	(2) Pre-determined controls	(3) Double Lasso Selection
Panel A: Direct effect			
Wedge follow up	−0.103***	−0.082	−0.104***
	(0.035)	(0.050)	(0.033)
Correct guess	0.053	0.084	0.064
	(0.057)	(0.055)	(0.053)
Distance from correct guess	5.133**	5.895**	4.539**
	(2.296)	(2.979)	(2.030)
Absolute distance from correct guess	−3.998**	−5.033***	−4.235***
	(1.686)	(1.679)	(1.542)
N	218	217	217
Panel B: Spillover effect			
Wedge follow up	−0.057**	−0.055*	−0.039*
	(0.029)	(0.028)	(0.022)
Correct guess	0.052	0.050	0.051
	(0.047)	(0.056)	(0.044)
Distance from correct guess	0.190	0.354	−0.760
	(1.972)	(1.792)	(1.625)
Absolute distance from correct guess	−0.082	−0.525	−0.422
	(1.109)	(1.161)	(0.921)
N	400	400	400

Notes: Model (1) shows the coefficient from the reduced form regression using only indicators for strata fixed effects as covariates. Model (2) shows the coefficient from a similar regression, but also including pre-determined controls. Model (3) shows the coefficient from a PDS-lasso regression, utilizing a post double lasso selection procedure in order to determine which covariates to include. Robust standard errors clustered at class level in parenthesis below each coefficient. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

The first row of Panel A suggests that being made eligible for treatment is associated with a decrease in guessed proportion below median. The measure *wedge* indicates the difference between the student's guessed proportion of classmates being below median and the actual proportion below the median. A negative estimated coefficient implies that students eligible for treatment on average guess that fewer of their peers are below the median. Model (1) and (3) are statistically significant at a 1 % level, and while model (2) is not significant at any conventional level, a *p-value* of 0.106 tells us it is very close to the 10 % level. It is important to note that the students were not tasked with guessing how many of their peers were below the median in the follow up survey, but that they were tasked with guessing how many of their peers reported being below median *in the main survey*.

The third row of Panel A suggests that students made eligible for treatment guess a higher median than the actual median in their class. The findings in row 4 suggest that the absolute difference between guessed median and actual median is decreasing for eligible students. An explanation for this might be that students were underestimating *ex ante*. The treatment nudged these students to reconsider their initial guesses, and on average, increase their guessed median. Consolidated, the results from row 1, 3 and 4 suggest that while students are increasing their relative distance to the correct median (that is, they are overestimating more), the net effect of reducing the prior underestimation is such that the absolute distance is reduced. The total effect seems to be that students eligible for treatment are better at guessing than their non-treated peers.

We also see some spillover effects in the first row of Panel B. Model (1) and (2) report a coefficient at about 50 % the magnitude of the direct effect, while model (3) has estimates a somewhat weaker relationship. All models are statistically significant at conventional levels. The estimated coefficient implies that there are some dynamic between the students that allow for the treatment to also affect some of the non-eligible students. More specifically, students having peers below median who are made eligible for treatment in their class are more inclined to guess a higher proportion above median.

6.2 Instrumental variable estimation

Moving on to the instrumental variable estimation, we now consider the *local average treatment effect*. Table 6.3 shows the beta coefficient from estimating equations 5.1 and 5.3 as in a two-stage least squares regression. Due to the mechanical aspect of our analysis, we do not find it necessary to show the first stage regression. By assumption our instrument is highly relevant, and any additional information from the first stage is not key in analyzing our data. However, all first-stage regressions are reported in Appendix 2.

The first row of Panel A in Table 6.3 suggests that being treated is associated with an increase in time planned on homework in the future. This suggestion is further reinforced by the second row which implies that treated students also have a higher probability of planning more homework than the median relative to their non-treated peers. The coefficients seems to be of about the same magnitude as their ITT counterparts, with the exception of model (3) which reports a slightly stronger effect. Considering that the LATE is a local measure for the effect on the treated and the ITT effect only considers the *eligibility* for treatment, it would be reasonable to expect such a difference.

Table 6.3: Direct treatment effects from the main survey

	(1) No controls	(2) Pre-determined controls	(3) Double Lasso Selection
Panel A: Direct effect			
Time planned homework	8.329***	7.964***	9.157***
	(2.257)	(2.491)	(2.031)
Planned homework above median	0.158***	0.229***	0.165***
	(0.056)	(0.054)	(0.056)
N	287	286	286
Panel B: Spillover effect			
Time planned homework	−0.900	−2.125	−2.873
	(2.965)	(2.826)	(2.802)
Planned homework above median	−0.000	−0.022	−0.028
	(0.029)	(0.028)	(0.029)
N	504	504	504

Notes: Model (1) shows the coefficient from an IV-regression using only indicators for strata fixed effects as covariates. Model (2) shows the coefficient from a similar IV-regression, but also including pre-determined controls. Model (3) shows the coefficient from an IV-lasso regression, utilizing a post double lasso selection procedure in order to determine which covariates to include. Robust standard errors clustered at class level in parenthesis below each coefficient. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

The difference in estimated coefficients between the three models are mainly driven by the covariates. We see that model (1) and (2) are quite similar, with (2) reporting a slightly lower coefficient. Model (3) differs somewhat more, suggesting about 10 % stronger effect in *time planned on homework*. For probability of time planned to be above the median, the estimate from model (2) is higher than the two other. Significance seems stable across all three models.

Overall, these estimations imply that informing students about the median time their peers spend on homework is sufficient to nudge them to plan more study time for themselves. Furthermore, the effect seems to be strong enough to substantially increase the proportion of below-median students who plan to do more homework than the median.

Table 6.4 is comparable to table 6.2, however with estimated LATEs instead of ITT effects. The estimated coefficients in the first row of Panel A suggests a positive effect on students ability to correctly guess their peers answers. The magnitude of the coefficients are quite similar to those from the ITT estimation. Model (2) is significant at a 5 % level, while model (1) and (3) exhibit an even higher significance level of 1 %.

The second row of Table 6.4 suggests a positive effect on students ability to perfectly guess their peers answers on time use. The coefficient of model (2) implies that students whom receive treatment are 10.9 percentage points more likely to correctly guess their peers answers perfectly. The estimate is significant at a 5 % level. Model (1) and (3) do not suggest any significant treatment effects.

The third row further confirms the relationship suggested in 6.2, however with the main difference that model (3) yields a non-significant estimate. The fourth row follows in the same fashion, with model (1) and (2) having the same interpretation as in table 6.2, while model (3) is non-significant.

We also find some evidence for spillover effects in the local treatment effects. The first row of model (1) and (2) of Panel B suggests that students whom are not treated but who are in the same class as someone treated, guess that 7.3 (5.7) percentage points more of their peers report spending more time than the median in their class on homework. Model (1) and (2) estimate a statistically significant relationship at the 5 % level, while model (3) is only significant at the 10 % level with a weaker estimated relationship.

Table 6.4: Direct and indirect treatment effects from the follow-up survey

	(1) No controls	(2) Pre-determined controls	(3) Double Lasso Selection
Panel A: Direct effect			
Wedge follow up	−0.109***	−0.091**	−0.097***
	(0.034)	(0.042)	(0.037)
Correct guess	0.066	0.109**	0.052
	(0.054)	(0.046)	(0.051)
Distance from correct guess	4.930**	5.939**	2.890
	(2.150)	(2.428)	(2.091)
Absolute distance from correct guess	−3.796**	−4.977***	−2.148
	(1.587)	(1.470)	(1.511)
N	218	217	217
Panel B: Spillover effect			
Wedge follow up	−0.073**	−0.057**	−0.059**
	(0.030)	(0.025)	(0.028)
Correct guess	0.077	0.083	0.069
	(0.049)	(0.054)	(0.054)
Distance from correct guess	1.266	0.182	0.211
	(2.163)	(1.550)	(1.695)
Absolute distance from correct guess	−0.482	−1.008	−1.234
	(1.148)	(1.065)	(1.244)
N	400	400	400

Notes: Model (1) shows the coefficient from an IV-regression using only indicators for strata fixed effects as covariates. Model (2) shows the coefficient from a similar IV-regression, but also including pre-determined controls. Model (3) shows the coefficient from an IV-lasso regression, utilizing a post double lasso selection procedure in order to determine which covariates to include. Robust standard errors clustered at class level in parenthesis below each coefficient. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Consolidating our findings, we see some evidence for positive treatment effect, as well as some spillover effects. The results remain significant across several model specifications.

6.3 Heterogeneity analysis

The following section presents the results from a heterogeneity analysis. Based on the existing literature, we have reasons to believe that different sub-samples might respond differently to the treatment. In order to test for these differences, we re-estimate our model including an interaction term between the suspected heterogeneity indicator and the treatment variable. In essence we estimate model 5.4. We perform separate estimations across all characteristics we suspect might be subject to heterogeneity issues. Table 6.5 summarizes the coefficients of the interaction terms from the estimations.

Table 6.5: Heterogeneity analysis

	Math grade > 4	Norwegian grade > 4	Gender Female	Utility homework Low	Popularity homework Low
Panel A: Direct effect					
Wedge follow up	–0.006	–0.020	–0.023	0.205**	0.177**
	(0.068)	(0.069)	(0.061)	(0.096)	(0.077)
Correct guess	0.093	–0.010	0.180*	–0.231*	–0.042
	(0.102)	(0.108)	(0.107)	(0.134)	(0.133)
Distance from correct guess	1.960	3.844	2.049	–14.601***	–3.262
	(3.748)	(4.256)	(3.442)	(4.129)	(3.853)
Absolute distance from correct guess	–8.389***	–5.995**	–1.972	12.258***	1.090
	(2.557)	(3.021)	(2.624)	(3.758)	(3.223)
N	218	218	217	218	218
Panel B: Spillover effect					
Wedge follow up	0.063	–0.062	0.038	–0.024	0.043
	(0.051)	(0.041)	(0.049)	(0.047)	(0.052)
Correct guess	0.054	–0.012	–0.117	–0.073	–0.204**
	(0.087)	(0.084)	(0.081)	(0.082)	(0.099)
Distance from correct guess	–0.582	5.017*	–4.526	3.135	1.774
	(2.926)	(2.700)	(3.040)	(3.293)	(3.439)
Absolute distance from correct guess	–3.606*	–1.862	3.137	1.189	5.698**
	(1.848)	(2.185)	(2.119)	(2.551)	(2.659)
N	400	400	400	400	400

Notes: Heterogeneity analysis by adding an interaction term to the IV-regression. The model controls for pre-specified covariates and is coinciding with Model (2) from the main analysis. Robust standard errors clustered at class level in paranthesis. Each cell is a unique regression. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

6.3.1 Grades

According to Azmat and Iriberri (2010), higher-achieving students might respond differently to information treatment than low-achieving students. The authors argue that due to the high-achieving students' relatively higher ability to implement new information in their decision making, they should see a stronger effect among the top performing students. The first two columns of Table 6.5 show the coefficient of the interaction term between treatment and a dummy for whether the student achieved a grade of 5 or higher on their latest math and Norwegian test, respectively. We notice that only two of the coefficients are significantly different from zero, suggesting that students that achieve a grade of 5 or higher in math or Norwegian have a lower distance from the correct guessed median relative to the students achieving a grade of 4 or lower. The estimated effect is about 1.4 times as strong for the math grade than for the Norwegian grade.

From Panel B we see some evidence that high achieving students in math guess a lower absolute distance from median, and high achieving students in Norwegian guess a higher relative distance from the median. Both estimates are significant at a 10 % level.

These results suggest some heterogeneity in the treatment effect, dependent on the subjects grades. There seems to be some correlation between stronger treatment effect and high achieving students. This could have further implications for how to most efficiently implement such an intervention on a larger scale, and deserves careful consideration.

6.3.2 Gender

The third column of Table 6.5 suggests that female respondents are relatively better at guessing correctly (second row of Panel A). This effect is significant at a 10 % significance level, and the estimated coefficient is relatively large compared to the main effects from Table 6.3.

As with achievement, this result also supports the notion of some heterogeneity, however the evidence is only apparent for one of the outcome variables and the significance is weak. We should be cautious in interpreting these results.

6.3.3 Utility from homework

Another interesting notion is that the students' preferences might affect how perceptible they are to treatment. One could argue that students whom derive low utility from schoolwork might underestimate the amount of time their peers spend on schoolwork. For these students we would expect to see a stronger treatment effect due to the fact that their guesses are already disproportionately lower than their peers.

From column 4 in Table 6.5 we see some evidence of this intuition. All the coefficients displayed support the intuition that a weaker treatment effect is associated with students who derive low utility from homework. The first row suggests they guess that a higher proportion of their peers are below median in time spent on homework. The second row suggests a lower proportion of correct guesses. The third and fourth row consolidated implies that these students are worse at guessing, and that they have a propensity for underestimating, relative to their peers. All the estimates remain statistically significant at conventional levels.

Similarly, we see some significant effects in column 5. Both row 1 in Panel A and, row 2 and 4 in Panel B suggest that students who believe spending time on homework contributes to lower popularity is associated with a weaker treatment and spillover effect than their peers. We see a 20.4 percentage points lower spillover effect on proportion of perfect guesses relative to their peers, and their absolute distance to correct guess is 5.698 minutes higher than their peers.

6.3.4 Network effects

We test different network effects by constructing a measure for the average centrality of the treated students in each class. This measure is then used to create sub-samples of students in classes with higher or lower than average centrality for below-median students. The reported estimates show the difference in effect for students whom have treated peers that are more or less central than average. The results of the estimation is presented in Table 6.6. The first column shows the estimated effect of having the below-median students have higher eigencentrality than average. The remaining columns show the estimated effect of having the below-median students have a higher than average total amount of friends (degree), in-friends (indegree), out-friends (outdegree) and reciprocal friends, respectively.

Since we are interested in contact between students in order to facilitate treatment transfer, we allow one-way reported friendship ties to be interpreted as communicative. The rational behind this is that any form of friendship requires reciprocal involvement. Thus, it is reasonable to consider the different measures for friendship ties as indicators of interaction between students. Still, we estimate models separately for each measure as it allows for more nuanced discussion.

Table 6.6: Network effects: Centrality of Panel A

	Eigencentrality	Total friends	In-friends	Out-friends	Reciprocal friends
Panel A: Direct effect					
Wedge follow up	0.082	0.004	−0.193	0.027	0.006
	(0.065)	(0.083)	(0.121)	(0.067)	(0.068)
Correct guess	−0.033	−0.103	−0.000	0.006	0.098
	(0.112)	(0.127)	(0.160)	(0.119)	(0.142)
Distance from correct guess	−3.994	3.304	3.636	5.050	6.772*
	(3.205)	(4.770)	(7.489)	(3.946)	(4.096)
Absolute distance from correct guess	2.334	−0.485	1.108	−3.482	−6.900**
	(2.648)	(3.478)	(4.610)	(3.055)	(2.997)
N	217	217	217	217	217
Panel B: Spillover effect					
Wedge follow up	−0.006	0.045	−0.113*	0.102*	0.037
	(0.040)	(0.069)	(0.067)	(0.060)	(0.068)
Correct guess	−0.094	0.017	0.064	0.017	−0.055
	(0.068)	(0.085)	(0.120)	(0.100)	(0.105)
Distance from correct guess	−2.600	1.972	2.954	6.186*	4.128
	(2.751)	(4.270)	(4.322)	(3.648)	(3.091)
Absolute distance from correct guess	3.517*	−1.472	−1.070	−1.848	3.095
	(2.107)	(2.564)	(3.042)	(2.924)	(2.164)
N	400	400	400	400	400

Notes: Heterogeneity analysis by adding an interaction term to the IV-regression. The interaction term is dependent on network characteristics of the below-median students. The model controls for pre-specified covariates and is coinciding with Model (2) from the main analysis. Robust standard errors clustered at class level in paranthesis. Each cell is a unique regression. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Since we are mainly interested in investigating the heterogeneity in spillover effects, we begin by considering Panel B.

The fourth row of column one suggests that high eigencentrality among the treated students is associated with a larger difference between believed median and correct median. This is a somewhat surprising result as it implies that treating students with high centrality, and thus high propensity for interacting and initiating spillover effects, is associated with a weaker spillover.

Column two, three and four show the estimated coefficient from estimations with the interaction term depending on different measures of friendship ties. The first row suggests a correlation between the treated students having many in-friends and a stronger spillover effect, while having many out-friends is associated with a weaker spillover effect. We see no effect for reciprocal friends. These results are interesting as the they suggests the opposite of what one would expect. It would be reasonable to argue that treated students with many out-friends would have a higher frequency of interaction, and thus would be more inclined to see a stronger treatment effect.

Due to the ambiguity of the estimated effects, we do not find enough convincing evidence to argue for clear heterogeneity in spillover effect conditional on centrality measures. Our results do, however, motivate further investigation of this issue, as there seems to be some mechanism at play that our data set and specification is not able to pick up.

In Panel A we see some evidence that suggests heterogeneity in the direct treatment effect. The estimated coefficient of row 3 and 4 suggests an association between the treated students having many reciprocal friends, and a stronger treatment effect. The mechanism of this relationship could depend on treated students exchanging information about the treatment with each other, thus amplifying the treatment effect.

As discussed earlier, we are interested in the effect of interaction between treated and non-treated students. Since we are not able to directly observe such interaction, we rely on self-reported friendship ties. Following the discussion of the direction of such ties, it is also interesting to investigate any heterogeneity dependent on the network characteristics of the non-treated students. High frequency of interaction between above-median students could lead to heterogeneous spillover effects. This could for example be due to an above-median student seeking social interaction with several treated students. This mechanism closely relates to the friendship paradox Jackson (2019). We consider this effect in Table 6.7

These estimations suggests some differential effects for above median students with many friends. We see this through the first row of column 2 in both Panel A and Panel B. The estimated coefficient suggests that a high number of friends among the treated above median students is associated with a lower wedge. In addition, row four of the same column suggests a correlation between overestimation and total number of friends. In Panel B we find some evidence for weaker spillover effects, while still maintaining some of the overestimation also found in Panel A.

Table 6.7: Network effects: Centrality of Panel B

	Eigencentrality	Total friends	In-friends	Out-friends	Reciprocal friends
Panel A: Direct effect					
Wedge follow up	0.073	−0.278***	−0.094	−0.104	−0.021
	(0.084)	(0.071)	(0.070)	(0.077)	(0.054)
Correct guess	−0.279*	0.018	0.117	−0.220*	0.028
	(0.158)	(0.112)	(0.110)	(0.119)	(0.105)
Distance from correct guess	−9.462*	10.715***	5.644	1.624	1.418
	(4.980)	(3.249)	(4.405)	(4.147)	(2.973)
Absolute distance from correct guess	9.325**	−2.783	−3.872	1.991	0.753
	(4.172)	(3.411)	(3.360)	(3.979)	(2.712)
N	217	217	217	217	217
Panel B: Spillover effect					
Wedge follow up	−0.043	−0.105*	0.070	−0.035	−0.006
	(0.071)	(0.062)	(0.078)	(0.058)	(0.050)
Correct guess	−0.269**	−0.061	−0.356***	−0.044	−0.038
	(0.137)	(0.116)	(0.117)	(0.126)	(0.102)
Distance from correct guess	−1.020	6.351*	−4.569	2.036	3.356
	(4.679)	(3.735)	(4.919)	(3.601)	(2.775)
Absolute distance from correct guess	7.322**	3.143	9.393***	4.657	1.674
	(3.283)	(2.617)	(3.093)	(3.156)	(2.474)
N	400	400	400	400	400

Notes: Heterogeneity analysis by adding an interaction term to the IV-regression. The interaction term is dependent on network characteristics of the above-median students. The model controls for pre-specified covariates and is coinciding with Model (2) from the main analysis. Robust standard errors clustered at class level in paranthesis. Each cell is a unique regression. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Column three of Panel B suggests high amount of in-friends among above median students in treated classes is associated with a substantially lower proportion of correct guesses and higher distance from correct guess.

Finally, we specify our analysis to consider the difference between students with more or less than one treated friend. The estimations are consolidated in Table 6.8. We do not see any clear signs of heterogeneity in the spillover effects. The only exception is a weakly significant estimate that suggests some association between high number of treated friends and higher distance from correct guess.

In Panel A, however, we find some more interesting insights. First, the coefficient reported in the third row of column one suggests that having a lot of treated friends is associated with a lower distance from correct guess. Next, we find some evidence of a correlation between having many treated out-friends and stronger treatment effect. One reasonable explanation of this is that students who interact more with other treated students are subject to an amplification of treatment.

Table 6.8: Network effects: Number of treated friends

	In-friends	Out-friends	Reciprocal friends
Panel A: Direct effect			
Wedge follow up	0.040	−0.103*	0.016
	(0.028)	(0.059)	(0.028)
Correct guess	0.033	0.267**	0.027
	(0.048)	(0.117)	(0.063)
Distance from correct guess	−2.414*	5.257	−2.154
	(1.345)	(4.125)	(1.666)
Absolute distance from correct guess	−0.938	−5.797*	−0.230
	(1.065)	(3.174)	(1.418)
N	214	217	214
Panel B: Spillover effect			
Wedge follow up	0.010	−0.015	0.026
	(0.016)	(0.057)	(0.019)
Correct guess	−0.030	−0.066	−0.028
	(0.035)	(0.088)	(0.039)
Distance from correct guess	−0.610	−0.544	−2.353
	(1.386)	(3.409)	(1.487)
Absolute distance from correct guess	1.900*	3.191	1.480
	(1.094)	(2.609)	(1.154)
N	392	400	392

Notes: Heterogeneity analysis by adding an interaction term to the IV-regression. The interaction term is dependent on the number of treated friends of the subjects. The model controls for pre-specified covariates and is coinciding with Model (2) from the main analysis. Robust standard errors clustered at class level in paranthesis. Each cell is a unique regression. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

In total, our heterogeneity analysis does not provide any clear insight into the importance of network characteristics on treatment and spillover effects. We do find some implications of a relationship between the number of friends and the magnitude of treatment, but the results are ambiguous and require more in depth investigation in order to provide any constructive insight.

7 Robustness check

7.1 Balance and bias

Imbalances in baseline characteristics between treatment and control group could be a potential threat to causality. The main concern is that any of the imbalanced covariates are predictive of the outcome variable, thus contributing to omitted variable bias. This section is dedicated to discussing the potential effects the baseline characteristics could have on the outcome. In the following section, we expand on this analysis and then test whether our intuition holds when controlling for the imbalanced covariates.

Recalling the results from table 4.1, we see that the first sign of imbalance between treatment and control group is evident in the *Female* variable. Our data suggests that being a female student is associated with better initial guesses regarding peers' time spent on homework (see Appendix A3.1). A difference in baseline distribution of female and male students could then have an impact on the estimated treatment effect. Since the initial wedge is lower for female students, one reasonable suspicion is that the treatment will have limited effect. Thus, if we do not control for this imbalance in our analysis, we would expect to see a negative bias in our treatment effect (Panel A), and a positive bias in our spillover effect (Panel B).

We do not find any evidence of differential initial beliefs across the other imbalanced baseline characteristics. However, there seems to be some differences in initial time planned on homework. Both *female, Norwegian grade* > 4 and *importance grades* > 3 are associated with higher time planned to homework. The relationships are significant at a 1 % level. This does reflect some underlying difference in attitude across different baseline characteristics.

Another main concern with imbalanced covariates is that observed imbalances also reflect imbalances in the unobserved characteristics. This is a far more pressing issue, as we have no means of controlling for of measuring these kinds of differences.

7.2 Attrition

We check for attrition and it's effects by constraining our estimations to the respondents who participated in the follow-up survey. Our concern is that there is differential attrition rates conditional on treatment status, which could potentially skew our analysis by contributing to violation of the independence assumption and the condition of randomly distributed treatment.

From Table 7.1 we see that the attrition rate was on average 24% for the below median students and 20% for the above median students. The estimated coefficient in column 3 tells us that there is a 7 percentage points lower attrition rate for students above median in the treated classes. This suggests that students subject to spillover of treatment effects are more likely to stay in the study relative to their untreated peers. We don't see any evidence for differential attrition for the below-median students.

Table 7.1: Attrition rate

	Mean C	Mean T	T-C	(se)	n.obs
Panel A: Below median					
Attrition rate	0.24	0.24	–0.01	(0.045)	287
Panel B: Above median					
Attrition rate	0.24	0.17	–0.07**	(0.033)	504

Notes: Attrition rate by regressing a dummy for participation in the follow-up survey on treatment status. Controlled for strata fixed effects and standard errors clustered at class level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

To investigate whether attrition might affect the distribution of baseline characteristics in our sample, we re-estimate our model for balance testing on a subsample consisting of only students whom respond to both surveys. The differences in baseline characteristics are summed up in table 7.2.

Relative to the full sample, we see that there are indeed a few more baseline characteristics that are significantly different between the treatment and control group. For Panel A, we see imbalance in the gender indicator and returns to studying Norwegian. For Panel B, we find the same imbalances as for the full sample with an addition of the number of friends. The differences are about the same, but the significance is lower. Interestingly, we do not see any evidence for differences in initial beliefs in the restricted sample, as we do in the full sample. The discrepancy between significant variables in the full sample and the restricted sample does indeed imply that attrition might prove an issue for the internal validity of our results.

Table 7.2: Baseline characteristics of subsample responding to follow-up survey

	Mean C	Mean T	T-C	(se)	n.obs
Panel A: Below median					
Female	0.30	0.44	0.15**	(0.059)	217
Time homework	14.95	15.05	1.79	(1.797)	218
Dislikes homework	0.84	0.77	−0.00	(0.046)	218
Grades					
Standardized grade math	0.01	−0.017	−0.02	(0.130)	218
Standardized grade Norwegian	−0.09	−0.15	−0.08	(0.153)	218
Returns to studying Norwegian	0.51	0.44	−0.12*	(0.060)	218
Returns to studying math	0.64	0.55	−0.09	(0.097)	218
Attitudes					
Importance grades > 3	0.90	0.84	−0.05	(0.040)	218
Pleasing parents > 3	0.65	0.63	0.01	(0.071)	218
Expectation of parents > 3	0.92	0.86	−0.07	(0.052)	218
Pleasing teachers > 3	0.47	0.49	−0.01	(0.082)	218
Expectations of teachers > 3	0.83	0.83	−0.02	(0.046)	218
Importance of popularity > 3	0.78	0.73	−0.05	(0.054)	218
Popularity schoolwork < 3	0.25	0.19	−0.06	(0.055)	218
Friendship					
Number of in friends	3.18	2.94	−0.42	(0.319)	215
Number of out friends	3.02	2.96	−0.17	(0.315)	215
Number of reciprocal friends	6.19	5.90	−0.60	(0.617)	215
Eigencentrality	0.45	0.51	0.04	(0.046)	215
Many friends outside class	0.50	0.36	−0.10	(0.061)	218
Beliefs					
Guessed median - class median main survey	−18.56	−17.30	−0.31	(1.742)	218
abs(guessed median - class median) main survey	19.59	19.93	0.81	(1.354)	218
Wedge main survey	0.34	0.32	−0.03	(0.026)	218
Guessed median = class median main survey	0.10	0.09	0.00	(0.042)	218
Panel B: Above median					
Female	0.66	0.57	−0.12***	(0.041)	400
Time homework	48.50	47.52	−0.08	(1.605)	400
Dislikes homework	0.62	0.58	−0.04	(0.041)	400
Grades					
Standardized grade math	0.12	0.15	−0.03	(0.073)	400
Standardized grade Norwegian	0.26	0.13	−0.20**	(0.076)	400
Returns to studying Norwegian	0.31	0.37	0.05	(0.044)	400
Returns to studying math	0.59	0.57	0.03	(0.046)	400
Attitudes					
Importance grades > 3	0.93	0.84	−0.08**	(0.030)	400
Pleasing parents > 3	0.72	0.78	0.09**	(0.038)	400
Expectation of parents > 3	0.92	0.92	−0.00	(0.024)	400
Pleasing teachers > 3	0.62	0.67	0.06	(0.039)	400
Expectations of teachers > 3	0.88	0.88	−0.01	(0.027)	400
Importance of popularity > 3	0.76	0.79	0.04	(0.027)	400
Popularity schoolwork < 3	0.16	0.21	0.05	(0.032)	400
Friendship					
Number of in friends	3.04	2.93	−0.28	(0.199)	392
Number of out friends	3.23	2.93	−0.48**	(0.221)	392
Number of reciprocal friends	6.27	5.86	−0.76*	(0.404)	392
Eigencentrality	0.47	0.52	0.03	(0.031)	392
Many friends outside class	0.39	0.37	0.00	(0.045)	400
Beliefs					
Guessed median - class median main survey	−4.04	−3.88	−0.31	(2.087)	400
abs(guessed median - class median) main survey	12.90	13.63	1.24	(1.019)	400
Wedge main survey	0.14	0.12	−0.04	(0.027)	400
Guessed median = class median main survey	0.27	0.27	0.01	(0.031)	400

Notes: Female is a manually coded variable based on the name of the student. The first and second columns show the mean value of the row variables for the control and treatment group, respectively. The third column shows the estimated coefficient from a regression of the baseline characteristic on treatment status, controlling for strata fixed effects. The fourth column includes robust standard errors, clustered at the class level (corresponding with randomization level). Each row includes a separate regression. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

We test for the impact of these imbalances by estimating a model without any controls and a model where we control for the imbalanced covariates, and compare them. Table 7.3 shows the estimated coefficients of treatment for the two different models. We see that the coefficient for *wedge* is significant for both panels and both models. Considering the other outcome variables, we see that *distance* and *absolute distance* both have equally significant effects. Moreover, model (2) seems to be estimating a somewhat stronger treatment effect. In Panel B, the two estimated effects are almost identical, both in magnitude and significance.

Table 7.3: IV estimation controlling for unbalanced subsample

	(1) No controls	(2) Controlling for imbalance
Panel A: Direct effect		
Wedge follow up	−0.109***	−0.108***
	(0.034)	(0.036)
Correct guess	0.066	0.050
	(0.054)	(0.057)
Distance from correct guess	4.930**	5.792**
	(2.150)	(2.331)
Absolute distance from correct guess	−3.796**	−4.576**
	(1.587)	(1.802)
N	218	214
Panel B: Spillover effect		
Wedge follow up	−0.073**	−0.072**
	(0.030)	(0.030)
Correct guess	0.077	0.070
	(0.049)	(0.046)
Distance from correct guess	1.266	1.690
	(2.163)	(2.024)
Absolute distance from correct guess	−0.482	−0.207
	(1.148)	(1.210)
N	400	392

Notes: Model (1) shows the coefficient from an IV-regression using only indicators for strata fixed effects as covariates. Model (2) shows the coefficient from a similar IV-regression, but also including controls for imbalanced baseline characteristics. Robust standard errors clustered at class level in parenthesis below each coefficient. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

While our discussion of the attrition rate raised some concern about the internal validity of our analysis, we do not see any evidence of stark differences between a model estimated with and without controls for the imbalanced covariates. Our main outcome measures remains stable across both models, which suggests that differential attrition does not affect the analysis in any noticeable way.

7.3 Issues with randomization

For our main analysis, we use an IV approach. This is to avoid issues with non-compliance and treatment dilution. However, we also have a variable in our data set which indicates what schools were experiencing issues with the implementation of the treatment. In the absence of non-compliance, the estimated ITT effect corresponds with the LATE. The following section will present our main results from the IV analysis and compare them to a alternative model where we restrict our sample to only the classes with successful implementation of treatment.

From Table 7.4 we see that the differences between the two models are minuscule for all outcomes except the wedge measure in Panel B. Here, both the significance and the estimated effect is substantially smaller for the restricted sample. For all other variables, both the estimate and the significance support the same conclusion regarding the effect of treatment. The differences we see in the estimates could be the result of a smaller sample in one of the models.

Table 7.4: Difference between ITT on successful randomization and IV on full sample

	(1) Issue = 0	(2) IV
Panel A: Direct effect		
Wedge follow up	–0.095*	–0.091**
	(0.050)	(0.042)
Correct guess	0.100*	0.109**
	(0.060)	(0.046)
Distance from correct guess	5.711*	5.939**
	(3.274)	(2.428)
Absolute distance from correct guess	–5.515***	–4.977***
	(1.805)	(1.470)
N	191	217
Panel B: Spillover effect		
Wedge follow up	–0.047	–0.057**
	(0.031)	(0.025)
Correct guess	0.030	0.083
	(0.059)	(0.054)
Distance from correct guess	0.257	0.182
	(1.833)	(1.550)
Absolute distance from correct guess	0.011	–1.008
	(1.236)	(1.065)
N	330	400

Notes: Model (1) shows the coefficient from an reduced form regression restricted to the school where there are no confirmed issues with randomization. Model (2) shows the coefficient from a IV-regression on the whole dataset, corresponding with the main model. Both model controls for predetermined baseline characteristics, as well as strata fixed effects. Robust standard errors clustered at class level in parenthesis below each coefficient. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

8 Discussion

8.1 Potential mechanism

Our thesis depends on students misperceptions about peers' time spent on homework. While there are many possible explanations for this misperception, one reasonable intuition is related to misreporting among students. For example, if a student thinks that everyone else in their peer group is spending a small amount of time doing homework, they will naturally believe that doing less homework is socially desirable. However, such beliefs might be incorrect, and the stigma associated with this attitude could lead to incorrect beliefs. If the students believe that doing more homework is stigmatized, they might be reluctant to talk about it and reveal their private views to others. If most students act this way, they might all end up believing that their private views are only shared by at best a small minority. Furthermore, this leads to a preference falsification, in which the students reported preferences are affected by social acceptability, providing a biased summary of the class' view on homework. This could potentially update the student's beliefs about the share of people with negative views towards homework and make the students reluctant to talk about and reveal their private views to their peers.

Our analysis suggests some evidence of the above intuition. We find that the baseline difference in guessed proportion and the actual proportion of peers below median time spent on homework is high. Since students cannot observe the actual time spent on homework among their peers, they are dependent on the information their peers provide them. Thus, it is reasonable to attribute some of the differences between the actual and the perceived situation to misreporting. Our analysis also suggests positive spillover effects, implying that updating some students' beliefs with correct information about their peers' time spent on homework reduces the misreporting among students.

8.2 Limitations of the data set

As mentioned in Section 4.1, our study is based on self-reported data. The justification of this is that only the students themselves can report how much time they plan to allocate to different activities. In addition, most interactions in friendships occurs outside of public view, or at least away from the attention of adults. Thus, it seems necessary to rely on the students to report information about planned allocation of time and friendship ties. However, being in their adolescence the students are not necessarily reliable reporters of their experiences in group relationships or regarding time spent on different activities. Misreporting could lead to unreliable datasets in relation to the actual situation, however, this issue is equally present in the control group and the treatment group, and thus does not affect internal validity.

Most of the questions the students are presented with in the surveys have pre-determined options in which the students have to choose from. However, some of the questions allows for the students to write the answer themselves, e.g., writing the date of their latest test in a given subject. These open questions present some problems, as we see a tendency of misreporting among the students. In order to get a representative data set, we depend on the students writing a valid answer. An aspect that further increases this concern, is the monetary price connected to the surveys. By completing the follow-up survey for instance, the students become part of a lottery in which three of the students in each class are rewarded 200 NOK. This might lead some students to quickly going through the survey, not paying attention to what they answer, with the sole purpose of being part of the lottery. Moreover, this could lead to a misrepresentation of how many of the students actually improved their beliefs after being treated, which in turn could lead to falsified spillover effects.

A second potential problem is the fact that the female variable is manually coded based on the first name of the student. We must take into account that the proportion of girls might be incorrect. Both because there might be gender neutral names placed in the wrong category, or simply that some girls have "boys'" names or vice versa. The uncertainty regarding this baseline characteristic might be problematic, and it is important to interpret these results with caution. Generally, it is hard to confirm that the effects we find are completely valid, as we do not have the true distribution of girls and boys among the students.

A third problem, regarding the spillover effects, is identifying the peers with the strongest source of influence. In our study, we assume that the students are affected by the peers they list as their friends. However, one could argue that adolescents could be more influenced by those whom they want to be friends with, or groups to which they aspire for membership. If the latter is true, even though the students reported friends are treated, we might not observe a spillover effect as expected if the ones they want to be friends with is not treated, or we might find a spillover effect in cases where a student's reported friends are not treated.

All of the above limitations are mainly a concern for external validity. Since the issues are present for both the treatment and control group, we would expect our analysis to return credible estimates applicable to our sample.

In our analysis we distinguish between in-friends, out-friends, total friends and reciprocal friends. In-friends is the number of peers who have reported a student as their friend, out-friends is the number of peers a student report as their friends, total friends is the total of the two, excluding any overlapping friendship ties, and reciprocal friends is the number of friendship ties where two students have reported each other. There is some uncertainty related to the dynamics of the different terms. Considering the nature of friendship ties, it is reasonable to believe that any one of these measures imply interaction between two students, regardless of being directed or undirected. The intuition behind this is that it is nearly impossible to consider a one-way friendship. Thus, the interpretation of the different measures could become somewhat complex, and we should be cautious when interpreting our analysis.

8.3 Limitations of the estimation strategy

As explained in subsection 5.1, there are three assumptions that must be satisfied in order for the IV-estimation to be meaningful, the relevance assumption, exclusion restriction, and the independence assumption.

The independence assumption is trivially satisfied because randomization is expected to lead to equally distributed confounders across assignment groups. However, we can never guarantee that the instrument itself is not correlated with the unobserved random effects captured in the regression model. If it is, the independence assumption would be violated, and our results would be biased.

One indirect way of assessing this assumption is to look at whether there is imbalance in measured covariates across the levels of the proposed instrument. Imbalance in measured covariates can in principle be eliminated by adjusting for them in the analysis, but such imbalances can be suggestive of imbalance across unmeasured variables. As shown in subsection 4.3.1, we find some differences between the treatment and the control group in both panels. Due to the slight imbalances we find in our dataset, we have to be careful when interpreting our results, as this might lead to the test returning a statistically significant effect when there in fact is no such effect, or non-significant results when there in fact is significant effects.

Furthermore, as IV estimations represent the effect of our variable of interest on individuals who react to the instrument, it is important to know who these individuals are in order to draw meaningful policy implications. When looking at the LATE in a randomized experiment, we have to consider the characteristics among the compliers compared to the population. Compliers may be composed of different sub-populations with different treatment effects, and as shown in subsection 6.3, we observe some heterogenous effects. Thus, we have to be cautious when interpreting and extrapolating the results.

8.3.1 Differences in estimated models

Throughout the analysis we have relied on three different specification in order to estimate our models. The main difference between them relates to the included covariates. Model (1) does not include any covariates. Under the strict assumptions for a successful randomized control trial, this specification should give us unbiased estimates of the treatment effect. However, as discussed earlier, there might be some predetermined characteristics that are predictive of the outcome variable. Including these in the analysis should increase the precision of our estimates. Thus, model (2) provides us with estimates that control for all pre-determined characteristics we believe *ex-ante* are affecting the students beliefs about time spent on homework. However, this model is susceptible to crowding out effects, especially for smaller sample sizes.

The final model was estimated using a post double lasso selection procedure. This ensures that all, if any, included covariates have a strong mechanical relation to the outcome variable. This method effectively removes our influence over the analysis, providing us with a neutral model without any subjective biases. We partial out the identifiers for strata fixed effect, but allow the selection procedure to decide which other covariates to include.

The issue of which model is the most reliable and precise, is open for discussion. With a lack of dramatic differences between the three models, it is difficult for us to argue for one model's superiority. Instead we rely on the notion that if all models are supporting the same conclusion, we can safely argue for a successful analysis.

8.4 External validity

An important concern with our design regards the external validity of the findings. There are several important qualifications to the generality of the treatment effects we estimate.

Our treatment and spillover effects are estimated from the behavior of a particular sample of students. In our paper, we consider a sample of 10th graders in Norway. First, we focus on change in our main outcome variable, planned time spent on homework, upon treatment. Second, we include the additional information students might acquire after interacting with their peers. The peers we study are very close – often friends and in the same class, and they form their associations naturally and endogenously.

As previously mentioned, adolescents are in a period of great vulnerability to peer effects in addition to a strong desire to fit in. These attributes are not specific to adolescents in our school, but general attributes for adolescent. However, there are many attributes in which the 10th graders we study might differ from other students. First, the Norwegian culture might be quite different from other countries, impacting how the students interact and are influenced by each other. Second, there are institutional differences between countries, making it hard to transfer our findings directly to other schools and countries. Third, the Norwegian homework structure might differ from those in other countries. In addition to these attributes, there might also be differences between countries, and how well developed they are. For example, a lot of peer influence among Norwegian students today might happen through their phones and social media. Students in some countries, however, might not have access to these means of communication, affecting the peer influence part of the study.

Even though some of our effects could be transferable to students in other schools and countries, it is hard to fully confirm this without further studies on the differences between schools and countries, and the importance of different independent variables for each of these.

8.5 Implications of the COVID-19 epidemic

The lockdowns in response to Covid-19 have interrupted conventional schooling with nationwide school closures in Norway. During the previous year, students have alternated between being at school and having digital classes at home. Since most of our experiment was conducted while covid unfolded, and because we can't be certain when and if the different subject schools were at school or home, there is a concern that this could have impacted our results. The main survey was always conducted at school with a minimum of one meter between each student, ensuring the students' privacy when answering the questions. However, there are especially two consequences of the epidemic and students having classes at home that potentially could impact our results.

First, the tasks students perform might not be identical to the situation in which there is physical attendance at school. Students doing everything from home could lead to most tasks being perceived as either homework or schoolwork, making it harder to distinguish between the two. Furthermore, this could lead to incorrect beliefs about how much homework one plans to do, and how much homework one has done. Second, both national and regional infection prevention measures have led to a significant reduction in student's opportunity to socialize with others outside of school, and they haven't been exposed to their usual social groups or classrooms. In addition to having classes at home, this severely reduces the opportunities to be affected by the other students and could thus have an impact on our spillover effect.

8.6 Long term vs short term

We find that the information contained in our intervention increased planned time spend on homework among the participants, as well as improving their precision in guessing their peers' time spent on homework in the short run (about 2 weeks after intervention).

The magnitude of the effects we observe may reflect the relatively short time horizons over which outcomes are measured. In regard to the beliefs on how much time peers spent on homework, the outcome tests belief on schoolwork that was done during the previous week. This deviates from many other educational studies, which cover a semester or a year's worth of material. If beliefs decay over time, it may result in a reduction of the measured impact of the treatment over time. This could have methodical implications for future research as more frequent follow-up testing could allow for more precise estimation of long-term effects.

9 Conclusion

This master thesis studies the effect of providing 10^{th}-grade students with information about their peers time spent on homework. The data was collected from an ongoing experiment conducted by researchers at NHH. In total our data set consists of observations from 17 schools. The study is based on students' self reported study time and grades, as well as preferences towards homework and other activities.

Our results suggests that treatment both increases the future time planned on homework and improves the students precision in guessing their peers' time spent on homework. One hypothesis supported by the results is that treated students are underestimating *ex ante* and treatment nudges them to increase their estimates, resulting in more overestimation *ex post*. However, the total effect shows a positive development in overall precision. Our analysis also suggests an association between negative attitudes towards homework and a weaker treatment effect, and high achievement and stronger treatment effect.

We find some evidence of spillover effects, particularly regarding students' precision in guessing. Students not eligible for treatment in treated classes see a spillover effect in the same direction as the treatment effect, but with reduced magnitude.

All results remain relatively stable across different estimation specifications and sub-samples. We run robustness checks by including imbalanced baseline covariates and by controlling for attrition, without seeing any substantial difference in results. We also test for heterogeneous spillover effect conditional on centrality measures in the network, but do not find any compelling evidence.

The main policy implication of these findings relate to the efficiency of informational interventions in the school system. We see that a simple informative notice has the potential to change students intentions in the short term, and also influence their perceptions in the longer term. These types of interventions are relatively low-cost and low-impact, and could be implemented on a larger scale without substantial effort.

The focus of our analysis has been on the perceptions and beliefs of 10^{th}-graders in Norway. Our thesis contributes to the existing literature by examining the treatment effect and transmission of treatment effects through social networks. While our analysis has focused solely on the beliefs of the students, it would be interesting to further investigate the direct behavioral impact such an intervention could have.

It would also be interesting to investigate the long term effects of this intervention. Our analysis is limited to the duration of the study, but it would be interesting to follow the students for a longer period of time to see how their beliefs and attitudes develop. An extension of this is also to further develop the heterogeneity analysis in order to better understand the dynamics that determine the treatment effect.

Finally, while we did not find any clear evidence for differential network effects, it would be interesting to focus more on role of friendship ties in diffusion and amplification of treatment effect.

References

Akerlof, G. A. (1991). Procrastination and obedience. *The American Economic Review*, 81(2):1–19.

Altman, D. G. (1985). Comparability of randomised groups. *Journal of the Royal Statistical Society: Series D (The Statistician)*, 34(1):125–136.

Angrist, J. D. (2006). Instrumental variables methods in experimental criminological research: what, why and how. *Journal of Experimental Criminology*, 2(1):23–44.

Angrist, J. D., Imbens, G. W., and Rubin, D. B. (1996). Identification of causal effects using instrumental variables. *Journal of the American statistical Association*, 91(434):444–455.

Angrist, J. D. and Pischke, J.-S. (2008). *Mostly harmless econometrics: An empiricist's companion*. Princeton university press.

Angrist, J. D. and Pischke, J.-S. (2014). *Mastering'metrics: The path from cause to effect*. Princeton University Press.

Ariely, D. and Wertenbroch, K. (2002). Procrastination, deadlines, and performance: Self-control by precommitment. *Psychological science*, 13(3):219–224.

Avvisati, F., Gurgand, M., Guyon, N., and Maurin, E. (2013). Getting parents involved: A field experiment in deprived schools. *Review of Economic Studies*, 81(1):57–83.

Azmat, G. and Iriberri, N. (2010). The importance of relative performance feedback information: Evidence from a natural experiment using high school students. *Journal of Public Economics*, 94(7-8):435–452.

Azmat, G. and Iriberri, N. (2016). The provision of relative performance feedback: An analysis of performance and satisfaction. *Journal of Economics & Management Strategy*, 25(1):77–110.

Bandiera, O., Larcinese, V., and Rasul, I. (2015). Blissful ignorance? a natural experiment on the effect of feedback on students' performance. *Labour Economics*, 34:13–25.

Banerjee, A., Chandrasekhar, A. G., Duflo, E., and Jackson, M. O. (2013). The diffusion of microfinance. *Science*, 341(6144).

Belloni, A., Chernozhukov, V., and Hansen, C. (2014). High-dimensional methods and inference on structural and treatment effects. *Journal of Economic Perspectives*, 28(2):29–50.

Brown, B. B. (2004). *Adolescents' relationships with peers*. John Wiley & Sons Inc.

Burke, M. A. and Sass, T. R. (2008). Classroom peer effects and student achievement. working paper 08-5. *Federal Reserve Bank of Boston*.

Bursztyn, L. and Jensen, R. (2015). How does peer pressure affect educational investments? *The quarterly journal of economics*, 130(3):1329–1367.

Carrell, S. E. and Hoekstra, M. L. (2010). Externalities in the classroom: How children exposed to domestic violence affect everyone's kids. *American Economic Journal: Applied Economics*, 2(1):211–28.

Cohen, J. M. (1977). Sources of peer group homogeneity. *Sociology of education*, pages 227–241.

Cooper, H. (1989). *Homework*. Longman.

Cooper, H., Robinson, J. C., and Patall, E. A. (2006). Does homework improve academic achievement? a synthesis of research, 1987–2003. *Review of educational research*, 76(1):1–62.

Corno, L. (1994). Student volition and education: Outcomes, influences, and practices. In *Portions of this chapter were presented at the annual meeting of the American Psychological Assn in Toronto, Canada, Aug 1993.* Lawrence Erlbaum Associates, Inc.

Costrell, R. M. (1994). A simple model of educational standards. *The American Economic Review*, pages 956–971.

Eren, O. and Henderson, D. J. (2008). The impact of homework on student achievement. *The Econometrics Journal*, 11(2):326–348.

Ersoy, F. (2019). Effects of perceived productivity on study effort: Evidence from a field experiment. *Available at SSRN 3253978*.

Falk, A. and Ichino, A. (2006). Clean evidence on peer effects. *Journal of labor economics*, 24(1):39–57.

Figlio, D., Giuliano, P., Özek, U., and Sapienza, P. (2019). Long-term orientation and educational performance. *American Economic Journal: Economic Policy*, 11(4):272–309.

Fryer Jr, R. G. (2016). Information, non-financial incentives, and student achievement: Evidence from a text messaging experiment. *Journal of Public Economics*, 144:109–121.

Garlick, R. (2013). Academic peer effects with different group assignment rules: Residential tracking versus random assignment. Technical report, Working paper, World Bank Development Research Group, Washington, DC.

Gerber, A. S., Arceneaux, K., Boudreau, C., Dowling, C., Hillygus, S. D., Palfrey, T., Biggers, D. R., and Hendry, D. J. (2014). Reporting guidelines for experimental research: A report from the experimental research section standards committee. *Journal of Experimental Political Science*, pages 81–98.

Gneezy, U., List, J. A., Livingston, J. A., Qin, X., Sadoff, S., and Xu, Y. (2019). Measuring success in education: the role of effort on the test itself. *American Economic Review: Insights*, 1(3):291–308.

Hansen, B. B. and Bowers, J. (2008). Covariate balance in simple, stratified and clustered comparative studies. *Statistical Science*, pages 219–236.

Hirshleifer, S. et al. (2015). Incentives for effort or outputs? a field experiment to improve student performance. *Unpublished manuscript. Cambridge, MA: Abdul Latif Jameel Poverty Action Lab (J-PAL).*

Hoxby, C. M. and Weingarth, G. (2005). Taking race out of the equation: School reassignment and the structure of peer effects. Technical report, Citeseer.

Imberman, S. A., Kugler, A. D., and Sacerdote, B. I. (2012). Katrina's children: Evidence on the structure of peer effects from hurricane evacuees. *American Economic Review*, 102(5):2048–82.

Jackson, M. O. (2019). The friendship paradox and systematic biases in perceptions and social norms. *Journal of Political Economy*, 127(2):777–818.

Kandel, D. B. (1978). Homophily, selection, and socialization in adolescent friendships. *American journal of Sociology*, 84(2):427–436.

Larson, R. and Richards, M. H. (1991). Daily companionship in late childhood and early adolescence: Changing developmental contexts. *Child development*, 62(2):284–300.

Metcalfe, R., Burgess, S., and Proud, S. (2019). Students' effort and educational achievement: Using the timing of the world cup to vary the value of leisure. *Journal of Public Economics*, 172:111–126.

Moffitt, R. A. (2000). Policy interventions, low-level equilibria and social interactions. In *SOCIAL DYNAMICS*, pages 45–82. MIT Press.

Mutz, D. C., Pemantle, R., and Pham, P. (2019). The perils of balance testing in experimental design: Messy analyses of clean data. *The American Statistician*, 73(1):32–42.

Nurmi, J.-E. (2004). Socialization and self-development. *Handbook of adolescent psychology*, 2:85–124.

Pop-Eleches, C. and Urquiola, M. (2011). Going to a better school: Effects and behavioral responses. nber working paper no. 16886. *National Bureau of Economic Research*.

Roberts, C. and Torgerson, D. J. (1999). Baseline imbalance in randomised controlled trials. *Bmj*, 319(7203):185.

Rury, D. and Carrell, S. (2020). Knowing what it takes: The effect of information about returns to studying on study effort and achievement. In *July. http://conference. iza. org/conference_files/edu_ 2020/rury_ d29540. pdf.*

Schulz, K. F., Altman, D. G., and Moher, D. (2010). Consort 2010 statement: updated guidelines for reporting parallel group randomized trials. *Annals of internal medicine*, 152(11):726–732.

Stinebrickner, R. and Stinebrickner, T. R. (2003). Working during school and academic performance. *Journal of labor Economics*, 21(2):473–491.

Stinebrickner, R. and Stinebrickner, T. R. (2004). Time-use and college outcomes. *Journal of Econometrics*, 121(1-2):243–269.

Stinebrickner, R. and Stinebrickner, T. R. (2008). The causal effect of studying on academic performance. *The BE Journal of Economic Analysis & Policy*, 8(1).

Thune, T., Reisegg, O., and Askheim, S. (2019). Skole og utdanning i norge.

Urberg, K. A., Değirmencioğlu, S. M., and Pilgrim, C. (1997). Close friend and group influence on adolescent cigarette smoking and alcohol use. *Developmental psychology*, 33(5):834.

Urminsky, O., Hansen, C., and Chernozhukov, V. (2016). Using double-lasso regression for principled variable selection. *Available at SSRN 2733374*.

Zimmerman, B. J., Bonner, S., and Kovach, R. (1996). *Developing self-regulated learners: Beyond achievement to self-efficacy.* American Psychological Association.

Appendix

A1 Survey

A1.1 Main survey

Figure A1.1: Main survey

Background

Introduksjon
Velkommen og tusen takk for din deltakelse! Dette er et forskningsprosjekt i regi av forskere ved Norges Handelshøyskole. Vi kommer til å stille deg noen spørsmål om skole og læringsmiljø.

Personvern
Alle svarene dine vil bli behandlet strengt fortrolig. Det vil ikke være mulig for lærere, foreldre eller andre elever å vite hvilke svar du gir.

Betaling
For å delta på denne undersøkelsen tjener du 50 kr. I tillegg vil du, som en del av selve studien, få delta i to ulike lotterier der det vil være mulig å tjene et begrenset pengebeløp. Totalbeløpet vil bli utbetalt i en lukket konvolutt innen noen få dager.

Hva er ditt deltakernummer?
Skriv inn deltakernummeret du nettopp fikk. Dette er viktig for å kunne gi deg riktig utbetaling.

	Click to write Column 1	
	Answer 2	Answer 3
Click to write Form Field 1		

https://nhh.eu.qualtrics.com/Q/EditSection/Blocks/Ajax/GetSurve...SurveyID=SV_20qjgMVe6Q5dYNL&ContextLibraryID=UR_9HOjiOUcTXs2tj7 Side 1 av 12

Time use

Tenk på en vanlig hverdag.

Hvis du ser bort fra tiden du er på skolen eller sover:
Hvor mye tid tror du at du vanligvis bruker på hver av følgende aktiviteter?

	Ikke noe tid	1-15 min	16-29 min	30-44 min	45-59 min	1 time eller mer
Egenpleie (dusjing, sminking, o.l.)	○	○	○	○	○	○
Snakke med foreldre, måltider, hjelpe til hjemme, passe småsøsken, o.l.	○	○	○	○	○	○
Lekser og skolearbeid etter skolen	○	○	○	○	○	○
Se på TV, serier, Youtube, o.l.	○	○	○	○	○	○
Sosiale medier (Facebook, TikTok, snapchat, chatting, o.l.)	○	○	○	○	○	○
Spille skjermspill (Playstation, PC-spill, mobilspill, o.l.)	○	○	○	○	○	○
Være med venner, kjæreste	○	○	○	○	○	○
Trening og organiserte fritidsaktiviteter (fotball, musikk, o.l.)	○	○	○	○	○	○
Lese bok, blader, aviser, o.l.	○	○	○	○	○	○
Jobbe for lønn utenfor hjemmet, frivillig arbeid	○	○	○	○	○	○

Du får nå en liste over de samme aktivitetene som i forrige spørsmål. Vi ønsker nå at du skal krysse av i hvilken grad du liker hver av disse aktivitetene.

	Misliker veldig	Misliker litt	Verken liker eller misliker	Liker litt	Liker veldig godt
Egenpleie (dusjing, sminking o.l.)	○	○	○	○	○
Snakke med foreldre/måltider/hjelpe til hjemme/passe småsøsken o.l.	○	○	○	○	○
Lekser og skolearbeid etter skolen	○	○	○	○	○
Se på TV/Serier/Youtube/osv.	○	○	○	○	○
Sosiale medier (Facebook, TikTok, snapchat, chatting, o.l.)	○	○	○	○	○
Spille skjermspill (Playstation, PC-spill, mobilspill, osv.)	○	○	○	○	○
Være med venner, kjæreste	○	○	○	○	○
Trening og organiserte fritidsaktiviteter (fotball, musikk, osv.)	○	○	○	○	○
Lese bok/blader/aviser	○	○	○	○	○
Jobbe for lønn utenfor hjemmet/frivillig arbeid	○	○	○	○	○

Grades

Du vil nå bli stilt noen spørsmål knyttet til dine karakterer.

Vi minner om at alle svarene vil bli behandlet strengt fortrolig. Det vil ikke være mulig for lærere, foreldre eller andre elever å vite hvilke svar du gir.

Hvilke karakterer fikk du i norsk og matte i slutten av 9. klasse?

	1	2	3	4	5	6
Norsk	○	○	○	○	○	○
Matte	○	○	○	○	○	○

Tenk på karakterene du fikk i slutten av 9. klasse.

Hvis du på hverdager hadde brukt 15 minutter mer på lekser og skolearbeid etter skolen forrige semester: Hvilke karakterer tror du da at du hadde fått i norsk og matte i slutten av 9. klasse?

	1	2	3	4	5	6
Norsk	○	○	○	○	○	○
Matte	○	○	○	○	○	○

I hvilken grad er du enig eller uenig med følgende utsagn om deg selv?

	Helt uenig	Delvis uenig	Hverken enig eller uenig	Delvis enig	Helt enig
Dette semesteret synes jeg det er viktig å få et karaktersnitt på minst 4.	○	○	○	○	○

I hvilken grad er du enig eller uenig med følgende utsagn om deg selv?

	Helt uenig	Delvis uenig	Hverken enig eller uenig	Delvis enig	Helt enig
Jeg liker å konkurrere	○	○	○	○	○
Jeg misliker å tape	○	○	○	○	○
Jeg liker å jobbe under press	○	○	○	○	○

Friends

Du vil nå bli stilt noen spørsmål knyttet til læringsmiljøet ditt, inkludert et spørsmål om hvilke elever du tilbringer mest tid med.

Vi minner om at alle svarene dine blir anonymisert og at det **ikke** vil være **mulig** for lærere, foreldre eller andre elever å vite hvilke svar du gir.

I hvilken grad er du enig eller uenig med følgende utsagn om deg selv?

	Helt uenig	Delvis uenig	Hverken enig eller uenig	Delvis enig	Helt enig
Jeg synes det er viktig å oppfylle **mine foreldres forventninger** til min innsats på skolen.	○	○	○	○	○
Jeg synes det er viktig å oppfylle					

lærernes forventninger til min innsats på skolen.	○	○	○	○	○

I hvilken grad er du enig eller uenig med følgende utsagn om dine foreldre og dine lærere?

	Helt uenig	Delvis uenig	Hverken enig eller uenig	Delvis enig	Helt enig
Mine foreldre synes det er viktig at jeg jobber hardt med lekser og skolearbeid etter skolen.	○	○	○	○	○
Mine lærere synes det er viktig at jeg jobber hardt med lekser og skolearbeid etter skolen.	○	○	○	○	○

Hvor ofte sjekker minst en av foreldrene dine leksene og skolearbeidet ditt?

	Aldri/nesten aldri	En gang i måneden	En gang i uken	Et par ganger i uken	Nesten hver dag
Antall ganger	○	○	○	○	○

I hvilken grad er du enig eller uenig med følgende utsagn om deg selv?

	Helt uenig	Delvis uenig	Hverken enig eller uenig	Delvis enig	Helt enig

Jeg synes det er viktig å ha mange venner i klassen.	○	○	○	○	○

Hva tenker du er viktig for å være populær og ha mange venner i klassen din?

	Reduserer populariteten mye	Reduserer populariteten litt	Har ingen betydning	Øker populariteten litt	Øker populariteten mye
Å være til å stole på	○	○	○	○	○
Å være flink i idrett	○	○	○	○	○
Å ha et karaktersnitt på minst 4	○	○	○	○	○
Å bruke i snitt minst 30 minutter på lekser og skolearbeid etter skolen på hverdager	○	○	○	○	○
Å få mange «likes» på sosiale medier	○	○	○	○	○
Å ha moteriktige klær	○	○	○	○	○

Vi minner om at alle svarene dine blir anonymisert og at det **ikke** vil være **mulig** for lærere, foreldre eller andre elever å vite hvilke svar du gir.

Her ønsker vi at du skal krysse av for hvilke elever du tilbringer mest tid med.

Listen inkluderer kun elever i klassen din, som deltar i studien i dag. Det kan derfor være det er noen du ikke finner. Du kan krysse av på så mange eller så få som passer for deg.

Navn 1

Navn 2

Navn 3

Jeg tilbringer ikke så mye tid med elevene som står på denne listen.

Er du mye sammen med 10-klassinger som ikke går i klassen din, men som går på skolen din?

Nei. Jeg er ikke mye med 10-klassinger som ikke går i klassen min, men som går på skolen min.

Ja. Jeg er mye med **noen få** 10-klassinger som ikke går i klassen min, men som går på skolen min.

Ja. Jeg er mye med **mange** 10-klassinger som ikke går i klassen min, men som går på skolen min.

Beliefs

Vi vil nå stille deg et spørsmål der du skal gjette hva elevene i klassen din tenker.

I tillegg til deltakeravgiften på 50 kr, vil alle som deltar i undersøkelsen fra klassen din få være med i et lotteri. Alle får et lodd til lotteriet hver, og på slutten av dagen vil vi trekke tre premier på 200 kr hver. Dersom du gjetter riktig på hva dine medelever tenker, får du et ekstra lodd i lotteriet. Gjetter du riktig dobler du dermed sjansen din for å få en av premiene på 200 kr.

Totalbeløpet, dvs. deltakeravgiften og evt. lotterigevinst, vil bli utbetalt i en lukket konvolutt innen noen få dager.

Fra klassen din deltar X elever i dag.

Vi ba dere tenke på en typisk hverdag: Hvor mange av dere X tror du svarte at de bruker minst 30 minutter på lekser og skolearbeid etter skolen?

Du får ett ekstra lodd i lotteriet dersom du svarer riktig.

***Antall** elever i klassen din som svarte at de bruker **minst 30 min** per dag.*

Final questions and info treatment

Du skal nå få litt (anonym) informasjon om hva elevene i klassen din har svart på et av spørsmålene. Merk at du vil bli bedt om å returnere til denne undersøkelsen, så vennligst ikke lukk dette vinduet.

Klikk på følgende link når det står på tavlen at det er klart: Klikk her

Når du har lest informasjonen som står der kan du returnere hit og gå videre.

På nettsiden du nettopp var innom stod det hvor mange i klassen din som har svart at de bruker minst 30 minutter per dag på lekser og skolearbeid etter skolen på en typisk hverdag.
Dette betyr at disse elevene bruker minst 30 minutter mer enn deg per dag.

På nettsiden du nettopp var innom stod det hvor mange i klassen din som har svart at de bruker minst 30 minutter per dag på lekser og skolearbeid etter skolen på en typisk hverdag.
Dette betyr at disse elevene bruker minst rundt 22 minutter mer enn deg per dag.

På nettsiden du nettopp var innom stod det hvor mange i klassen din som har svart at de bruker minst 30 minutter per dag på lekser og skolearbeid etter skolen på en typisk hverdag.
Dette betyr at disse elevene bruker minst rundt 7 minutter mer enn deg per dag.

Final question

Tenk på en typisk hverdag i tiden fremover til du er ferdig med 10. klasse.

Hvor mye tid planlegger du å bruke på lekser og skolearbeid etter skolen?

	Ikke noe tid	Mindre enn 15 min	16-29 min	30-44 min	45-59 min	1 time eller mer
Skolearbeid etter skolen	○	○	○	○	○	○

Du vil nå få delta i enda et lotteri (helt separat fra det forrige). I dette lotteriet vinner alle, men du kan påvirke hva du får i premie.

Nedenfor vil du bli bedt om å ta seks ulike valg. Hvert valg består av to alternativer. I hvert valg må du bestemme om du ønsker å få utbetalt et lite pengebeløp, eller om du vil motta motiverende teksmeldinger. **De motiverende**

tekstmeldingene skal hjelpe deg å nå målet du har satt deg om hvor mye tid du skal skal bruke på lekser og skolearbeid etter skolen i hverdagen.

Når undersøkelsen er ferdig **trekker vi tilfeldig ut et av valgene under som dermed bestemmer hva du får i premie fra dette lotteriet.**

For eksempel, dersom vi tilfeldig trekker ut "Valg 1" og du har svart at du foretrekker 5-ukers program med motiverende tekstmeldinger heller enn 0 kr, så vil du få et 5-ukers motiveringsprogram i tillegg til deltakeravgiften din på 50 kr, og premien fra det første lotteriet.

Alle pengebeløp, inkludert deltakeravgiften på 50 kr, vil bli utbetalt i en lukket konvolutt innen noen få dager.

Valg 1: Hva foretrekker du å vinne?

0 kr	5-ukers program med motiverende tekstmeldinger

Valg 2: Hva foretrekker du å vinne?

10 kr	4-ukers program med motiverende tekstmeldinger

Valg 3: Hva foretrekker du å vinne?

20 kr	3-ukers program med motiverende tekstmeldinger

Valg 4: Hva foretrekker du å vinne?

30 kr	2-ukers program med motiverende tekstmeldinger

Valg 5: Hva foretrekker du å vinne?

40 kr	1-ukes program med motiverende tekstmeldinger

Valg 6: Hva foretrekker du å vinne?

50 kr	0-ukers program med motiverende tekstmeldinger

Thank you for participating

Undersøkelsen er nå ferdig.

Tusen takk for at du deltok! Vi setter stor pris på din deltakelse!

LINK_TO_ALEX

Link

Powered by Qualtrics

A1.2 Follow-up survey

Figure A1.2: Follow-up survey

Qualtrics Survey Software 28.05.2021, 11:57

Background

Introduksjon
For noen uker siden deltok du på et forskningsprosjekt i regi av Norges Handelshøyskole. Prosjektet ble gjennomført på skolen din og vi er svært takknemlige for din deltakelse. Vi kontakter deg nå fordi vi håper du kan svare på en kort og enkel oppfølgingsundersøkelse. Denne oppfølgingsundersøkelsen er viktig for forskningsprosjektets suksess. Vi håper derfor at du har mulighet til å ta deg tid til å besvare denne. **Det vil kun ta et par minutter.** Undersøkelsen er frivillig, men svarene er viktige for prosjektet, og vi setter derfor stor pris på din deltakelse.

Personvern
Alle svarene dine vil bli behandlet strengt fortrolig. Det vil ikke være mulig for lærere, foreldre eller andre elever å vite hvilke svar du gir.

Betaling
Alle som deltar i undersøkelsen er med i et nytt lotteri. Alle får et lodd til lotteriet hver. Når alle har svart trekker vi **tre personer fra hver klasse som får 200 kr hver**. Vinnerene vil få beskjed per sms en stund etter at undersøkelsen er gjennomført slik at de kan motta utbetalingene.

Kontakt
Dersom du har noen spørsmål til undersøkelsen er det bare å ta kontakt med daglig ansvarlig, Ranveig Falch: ranveig.falch@nhh.no

R&F survey

Tenk på en vanlig skoledag de to siste skoleukene.

https://nhh.eu.qualtrics.com/Q/EditSection/Blocks/Ajax/GetSurveyP...tSurveyID=SV_0GSb01ylEOmixy5&ContextLibraryID=UR_9HOjiOUcTXs2tj7 Side 1 av 6

Hvis du ser bort fra tiden du var på skolen:
Hvor mye tid tror du at du vanligvis brukte per dag på lekser og skolearbeid?

Ikke noe tid

1 - 14 min

15 - 29 min

30 - 44 min

45 - 59 min

1t - 1t og 14 min

1t og 15 min - 1t og 29 min

1t og 30 min - 1t og 44 min

1t og 45 min - 1t og 59 min

2t - 2t og 14 min

2t og 15 min - 2t og 29 min

2t og 30 min - 2t og 44 min

2t og 45 min - 2t og 59 min

3t eller mer

Tenk på en vanlig skoledag de to siste skoleukene.

Hvis du ser bort fra tiden du var på skolen:
Hvor mye tid tror du at du vanligvis brukte per dag på hver av følgende aktiviteter?

	Ikke noe tid	Mindre enn 15 min	16-29 min	30-44 min	45-59 min	1 time eller mer
Egenpleie (dusjing, sminking o.l.).	○	○	○	○	○	○
Snakke med foreldre, måltider, hjelpe til hjemme, passe småsøsken o.l.	○	○	○	○	○	○

Se på TV, serier, Youtube o.l.	○	○	○	○	○	○
Sosiale medier (Facebook, TikTok, snapchat, chatting, o.l.).	○	○	○	○	○	○
Spille skjermspill (Playstation, PC-spill, mobilspill, o.l.).	○	○	○	○	○	○
Være med venner, kjæreste.	○	○	○	○	○	○
Trening og organiserte fritidsaktiviteter (fotball, musikk, o.l.).	○	○	○	○	○	○
Lese bok, blader, aviser o.l.	○	○	○	○	○	○
Jobbe for lønn utenfor hjemmet, frivillig arbeid.	○	○	○	○	○	○

Block 8

Du vil nå bli stilt noen spørsmål knyttet til dine karakterer.

Vi minner om at alle svarene vil bli behandlet strengt fortrolig. Det vil ikke være mulig for lærere, foreldre eller andre elever å vite hvilke svar du gir.

Vi ber deg tenke på <u>siste matteprøve</u> du tok (og som du har fått igjen karakter på).
Når du tok denne prøven?
(Skriv ca. dato. Dersom du ikke husker, ber vi deg gjette hva det kunne vært.)

Fyll inn dd.mm.åååå

Hvilken karakter fikk du på denne matteprøven?

	1	2	3	4	5	6
Karakter matteprøve	○	○	○	○	○	○

Vi ber deg tenke på siste norskprøve du tok (og som du har fått igjen karakter på).
Når du tok denne prøven?
(Skriv ca. dato. Dersom du ikke husker, ber vi deg gjette hva det kunne vært.)

Fyll inn dd.mm.åååå

Hvilken karakter fikk du på denne norskprøven?

	1	2	3	4	5	6
Karakter norskprøve	○	○	○	○	○	○

Block 7

Vi vil nå stille deg noen spørsmål der du skal gjette hvor mye tid elevene i klassen din bruker på lekser og skolearbeid utenom skoletid.

Alle som deltar i undersøkelsen fra klassen din får være med i et lotteri. Alle får ett lodd til lotteriet hver, og når alle har svart på undersøkelsen vil vi trekke tre premier på 200 kr hver. For hvert spørsmål du svarer riktig på, får du ett ekstra lodd i lotteriet. Gjetter du riktig øker du dermed sjansen din for å få en av premiene på 200 kr.

Vinnerne vil få beskjed per sms en stund etter at undersøkelsen er gjennomført slik at de kan motta utbetalingene.

Fra klassen din deltok ${e://Field/class_size} elever (inkludert deg) i vår undersøkelse for noen uker siden.

Hvis du <u>ser bort fra tiden på skolen</u>: Hvor mange av disse ${e://Field/class_size} **elevene tror du at på vanlige skoledager bruker:**

Ikke noe tid på lekser og skolearbeid.	0
Mellom 1 og 14 minutter på lekser og skolearbeid per dag.	0
Mellom 15 og 29 minutter på lekser og skolearbeid per dag.	0
Mellom 30 og 44 minutter på lekser og skolearbeid per dag.	0
Mellom 45 og 59 minutter på lekser og skolearbeid per dag.	0
Mer enn 1 time på lekser og skolearbeid per dag.	0
Total	0

Du får ett ekstra lodd i lotteriet for hvert riktige svar.
Husk: Summen av svarene dine kan ikke være mindre eller større enn ${e://Field/class_size}.

Block 4

I forbindelse med forskningsprosjektet, deltok du i en konkurranse. Har du og dine foreldrene snakket om den i etterkant?

Ja

Nei

Husker ikke

Dersom dere snakket om konkurransen, hva snakket dere om?

Powered by Qualtrics

A2 First Stage estimation

Table A2.1: Results from the first stage regressions

	$\hat{\phi}$	(se)	F-stat	N
Panel A: Direct effect				
Table 6.3 and 6.4	0.951***	(0.031)	74.26	217
Table 7.4 (1)	0.944***	(0.036)	1187.94	218
Table 7.4 (2)	0.941***	(0.027)	51898.25	214
Panel B: Spillover effect				
Table 6.3 and 6.4	0.853***	(0.067)	5.62	400
Table 7.4 (1)	0.859***	(0.066)	177.75	400
Table 7.4 (2)	0.850***	(0.069)	940.27	392

Notes: The table shows the estimated coefficient of the instrument from the first stage of the two stage least squares regression from the analysis. The model contains controls for pre-determined baseline characteristics and strata fixed effects. Robust standard errors, clustered at class level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

A3 Balance and bias

Table A3.1: The difference in initial beliefs conditional on baseline characteristics

	Time homework	Initial wedge
Female	9.94***	–0.05***
	(1.426)	(0.019)
Norwegian grade $>$ 4	5.69***	0.01
	(1.450)	(0.021)
Importance grades $>$ 3	8.80***	–0.01
	(2.375)	(0.028)
Expectation of parents $>$ 3	2.74	–0.02
	(2.691)	(0.025)

Notes: The table shows the estimated coefficients of a regression of the column-variable on the row-variable. Robust standard errors clustered at class level in parenthesis. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.