

Economics of Education: Peer Effects and Mindset

by

May Linn Auestad

Thesis submitted in fulfilment of
the requirements for the degree of
PHILOSOPHIAE DOCTOR
(PhD)



PhD programme in Social Sciences
UiS Business School
2020

University of Stavanger
NO-4036 Stavanger
NORWAY
www.uis.no

©2020 May Linn Auestad

ISBN: 978-82-7644-946-4

ISSN: 1890-1387

PhD: Thesis UiS No. 539

Acknowledgements

I am truly grateful for the opportunity to pursue a PhD within an area I care so deeply about. The financial support from the University of Stavanger, as well as the Norwegian Research Council and the project “*Usay: Shaping students learning expectations*” (grant number: 260407) are greatly acknowledged. Thinking back on the journey, I feel very blessed to have so many amazing and helpful people in my life, some of which I will try to express my gratitude towards below.

First, I want to start by thanking my main supervisor – Mari Rege – you are awesome! The way you push me to go far outside of my comfort zone is beyond anything I ever imagined possible. Despite maybe being the busiest person I know, you have taken the time to support and encourage me to pursue goals far beyond what I thought achievable. I will be eternally grateful for having had you as my supervisor!

Thank you, Ingeborg F. Solli, for always keeping your door open and making time for a chat! I also really appreciate you pushing me to write better, despite what it might have seemed like when we went through the feedback together afterwards. My thesis would not have been the same without you!

I also want to thank the last one of my supervisors – Eric Bettinger – for taking the time to sit down and explain when I had no idea of what I was doing. You make everything seem so easy! In addition, seeing you present have made me appreciate the skill of presenting that much more – you actually make it seem fun!

David S. Yeager – Thank you for important feedback on the pre-registration of the RCT, as well as for not saying “No!” to be a coauthor on the paper. In addition, just having witnessed you presenting and your thought process first hand is a learning process in and of itself – Thanks!

Marianne Bitler – Thank you for being my sponsor in the US as well as for the amazing feedback I got on the first paper in this thesis. The biggest influence you had, however, I do not think you are aware of yourself; Thank you so much for the life altering experience of not being that afraid to present! Your clear expectation in me wanting to present (which I most definitely did not want to do, but were too afraid to tell you), stands as the single most apparent transformation in my own beliefs in my ability to talk about my work in front of other people. Thank you so much for that!

Marianne Page – Thank you for your interest in spending time with me and getting to know me when I was in Davis. Our talks over breakfast, coffee and beer were very much appreciated. I have several times thought back to our discussions, which I believe have led me to see things from a somewhat different perspective than I would otherwise do. You really made me feel so welcome, and are one of the reasons for me missing Davis every once in a while. Thank you!

In addition, I also want to thank all of the other faculty members at UC Davis who took the time to sit down and talk with me when I was visiting, and for giving me feedback on the Brown Bag presentation. I also want to express my deepest gratitude to the graduate students at UC Davis for making me feel like I truly belonged there already from the start of my stay – you are another major reason why I occasionally miss being in Davis!

Aline Bütikofer – Thank you for helpful suggestions when working on my first paper, as well as useful tips when I was planning my trip to Davis. In addition, thanks for two amazing courses during my PhD – you helped make the transition back to academia very fun!

Venke F. Haaland – Thank you for having someone familiar to talk to in the start of my PhD. It is always nice to have a chat whenever you stop by, and I truly miss having you at the Business School!

Kjell Arne Brekke – Thank you for taking the time to read all of my work for my 90 percent seminar. I appreciate you making it a nice experience, as well as all the good comments and feedback I received.

Thank you to the people at the Centre for Learning Environment who included me in their discussions, which have helped a lot in understanding classroom settings. In addition, I would like to thank the employees at the participating school districts for important input and help facilitating the RCT, as well as all of the teachers who signed up for the project and gave me important insight into the teaching profession. Your reflections and feedback has been extremely important for my understanding of what is actually happening out there in schools.

Thank you to the people in Forsker Grand Prix, with a special thank you to Aud Berggraf Sæbø. The way you made me more comfortable in situations where I never ever imagined being, was truly mind-blowing. Thank you so much! I also want to express my gratitude towards the other participants, you made it feel more like a collaboration than a competition, and I really enjoyed spending time with you during the training sessions as well as at the finals at Stavangeren. Thanks!

Leigh Lauritzen – Thank you for taking the time to teach me some basic programming skills in a way in which I never felt like I was a burden. You opened my eyes to this completely new, and very fun, world. Thank you so much!

Thank you to the UiS Business School, our dean Ola Kvaløy, former dean Bjarte Ravndal and Christian Jensen, head of the economics and finance department. In addition, thank you to the people in the administration who always seem to find an answer to my sometime strange questions (regardless of the topic being work-related or not). Thank you also to the other colleagues at the Business School for nice lunch conversations as well as for always including the PhD students and making us feel like we belong.

I also want to express my deepest gratitude to the other PhD students for welcoming me with open arms when I started my PhD, as well as for always being there when the frustration of being a PhD student got a little bit too much to handle on my own. You have made this journey so fun! I really miss seeing all of you every day at the Business School. In addition, thank you Max and Andreas, for all of the discussions (work-related and maybe not so work-related) and feedback, as well as for understanding what I am talking about (when other people do not).

Thank you to all of my friends and family, for just being there whenever I need someone to talk to, as well as for giving me some escape from all things work-related. A special thanks to my parents – Liv Åse and Amund – for always making me believe that anything is possible! Thank you for always supporting me, regardless of the frustration you must have felt (which have become more apparent to me after I had my own children). You have always made me feel like every decision is my own and that you are going to support me regardless of what I end up doing. I do not think many people are as lucky as I am with the assigned parents. Thank you!

My two amazing children – Emil and Leon – Thank you for reminding me about the things that truly matter! Your enthusiasm and curiosity about everything really puts things into perspective and I am pretty sure the two of you are the primary reason for me getting through this PhD in time and with my sanity still there (or so I would argue). Thank you so much!

And last, but certainly not least, I want to thank my husband – Espen – who supports me in everything I do and still makes sure I do not lose sight of what is important – You truly are one of a kind!

May Linn Auestad
May 2020

Contents

Acknowledgements.....	iii
1 Introduction.....	1
2 The Norwegian educational system	5
3 Conceptual framework.....	9
3.1 Peer effects.....	10
3.2 Mindset	12
4 Empirical approach	15
4.1 Using observational data and experiments to uncover causal effects	15
5 Summary of papers	23
"The effect of low-achieving peers"	23
"Girls in STEM: The role of teacher mindset" (with Eric Bettinger, Mari Rege and Ingeborg F. Solli)	24
"Involving the teachers: Effects of a growth mindset intervention with teacher involvement" (with Mari Rege and David S. Yeager)	25
References.....	27
Paper I - III.....	37

1 Introduction

In a world where technology is becoming ever more important, children's best chance to thrive in school, and later, succeed in the labor market is by striving towards exploiting their full potential and acquire an applicable education. Evidence suggests that the future labor market will require more formal education, especially within science, technology, engineering and math, commonly referred to as the STEM fields (Freeman & Aspray, 1999; Petersen & Hyde, 2014). Additionally, mastering only the technical aspects of a job will no longer be sufficient; social and emotional skills will be essential as computers and robots take over routine tasks (Deming, 2017; Edin et al., 2017; Heckman et al., 2006; Heckman & Kautz, 2012; Kautz et al., 2014; OECD Publishing, 2013).

Furthermore, given that the advances in technology mean that the labor market is constantly changing, those who are able to practice lifelong learning are bound to experience some benefits. An obvious advantage in this regard is believing that intelligence and talents can be developed and changed through effort, perseverance, hard work and by using different strategies, commonly referred to as holding a growth mindset (Dweck, 2006; Mueller & Dweck, 1998; Yeager & Dweck, 2012).

Outside of the home, schools, and especially classrooms, are perhaps the most important arena for children developing the competencies they will need to flourish (Hamre & Pianta, 2005; Howes, Hamilton, & Matheson, 1994). The learning environment across or perhaps even within schools, however, can be rather different as there are several components that affect what happens in the classroom (Chetty et al., 2011; Rockoff, 2004). In this thesis I examine important aspects of two such components; other students, commonly referred to as peers, and teachers. My goal is to contribute to understanding how we can help more children

thrive in school, and consequently achieve more of their potential, using existing resources.

Inspired by the Bad Apple model of peer effects,¹ I start by investigating the effect of low-achieving peers in the first paper of this thesis. In particular, I look at the effect of exposure to at least one low-achieving student in the classroom on the performance of other students. Intuitively, we might expect that low-achievers require more of the teacher's attention, possibly taking important learning time away from their peers. If support is found for the Bad Apple model, this could imply that some students are not able to reach their full potential because of their low-achieving peers.

Empirical studies based on the Bad Apple model of peer effects do not seem to have reached a consensus (e.g. Carrell et al., 2010; Hanushek et al., 2002; Kristoffersen et al., 2015; Ruijs, 2017). Therefore, my contribution investigating the effect of low-achievers in Norwegian middle schools, where the focus on taking care of these students is more prominent than in many other countries, is important. In addition, by looking at the presence of at least one low-achiever, as opposed to the fraction of low-achievers, within the classroom, I also bring a new approach to the peer effect literature.

After concluding that students do not seem to be negatively affected by the presence of their low-achieving peers in Norwegian middle schools, I move on to examine the second component – teachers. In particular, I focus on the growth mindset, supposedly an important competence with regards to lifelong learning (Dweck, 2009). Intuitively, it is natural to assume that people holding more of a growth mindset set higher goals and are better able to reach them, because they are not afraid of making mistakes. Previous research suggests that holding more of a growth mindset has a positive effect on challenge-seeking behavior and

¹ Suggesting that one disruptive student harms everyone else (Hoxby & Weingarth, 2005; Sacerdote, 2011).

achievement (Andersen & Nielsen, 2016; Blackwell et al., 2007; Good et al., 2003; Paunesku et al., 2015; Yeager et al., 2019). In addition, some studies into how mindsets are developed suggest that teachers can, through their feedback and behavior, affect mindset, and thereby challenge-seeking behavior and achievement among their students (Dweck, 2006; Haimovitz & Dweck, 2017; Heggart, 2017; Mueller & Dweck, 1998). However, there is limited scientific evidence of this impact (Brooks & Goldstein, 2008).

In the second paper of this thesis, I and the coauthors address this gap in the literature using survey and registry data for all students starting high school in two of the largest counties in Norway in the school year 2017/2018. In addition, because of the gender gap in pursuing a career within the STEM fields, we focus on the subject most important for continuing towards a STEM degree, i.e. math, and perform gender specific analyses. We find that being exposed to a math teacher with more of a growth mindset in the last year of middle school – when important decisions about high school are being made – is beneficial for students' beliefs in their own abilities to learn when starting high school, especially with regards to math. In addition, we find a positive effect on participation in the most challenging math course in high school. We do not, however, find any effect on students' achievement in math or different effects based on students' gender; both girls and boys seem to benefit from being exposed to a teacher with more of a growth mindset.

Lastly, several studies have shown that exposing students to growth mindset interventions can increase the prevalence of such a mindset, and also have positive effects on student outcomes (Andersen & Nielsen, 2016; Bettinger et al., 2018; Blackwell et al., 2007; Paunesku et al., 2015; Yeager, Romero, et al., 2016; Yeager, Walton, et al., 2016). However, the interventions that are successful in increasing the prevalence of the growth mindset among students do not include teachers. Including the teachers could be beneficial for several reasons. For instance, it might help in scaling the interventions, or strengthen the positive effects by

teachers changing their practice (T. Kane et al., 2011; Mueller & Dweck, 1998).

In the third paper of this thesis, together with coauthors, I continue to focus on the role of teachers in developing a growth mindset among students, and address this gap in the literature. We do this by including teachers in the design of an intervention focused on increasing the prevalence of a growth mindset among students in high school. Thus conceivably making both teachers and students better at utilizing the learning opportunities that already exist. We test the intervention by performing a randomized controlled trial (RCT), and find positive effects on mindset as well as challenge-seeking behavior. The most important finding in this paper is that among students who are able to choose which math course to participate in, treated students are more likely to choose the more advanced course. In other words, they seem to be less afraid of challenge and more concerned with learning.

Understanding how different inputs, like peers and teachers, affect students' choices and learning is important for policy implications and in helping children reach their full potential (Heckman, 2000). The same goes for understanding how to develop social and emotional skills, like a growth mindset. For instance, the positive effect we find in papers two and three on the choice to participate in the more advanced math course could not only have implications for what students learn, but it could also affect their future career possibilities (Ma & Johnson, 2008; Sells, 1980).

The remaining part of my thesis starts with a section about the Norwegian educational system, which is the context of all three papers. I then present my conceptual framework, along with important aspects of the peer effects and mindset literatures. My empirical approach follows immediately after, where I start with a section about causation in general, before going deeper into the use of observational data and experiments in order to uncover causal effects. Lastly, I include brief summaries of the papers that follow in consecutive chapters.

2 The Norwegian educational system

In essence, formal education in Norway can be divided into four parts, as illustrated in Figure 1. First, children enroll in 1st grade at primary school (elementary school) in August the calendar year they turn six. Then, the year they turn thirteen (when starting 8th grade) they move on to lower secondary school (middle school). Together, the primary and lower secondary schools constitute the mandatory education. All children living in Norway for at least three months have both a right and a duty to participate in the mandatory education (the Ministry of Education and Research, 1998). Non-compliance with the enrollment regulation as well as repeating or skipping a grade is not common in Norway, meaning that children usually graduate from middle school with their age cohort in the calendar year they turn sixteen.

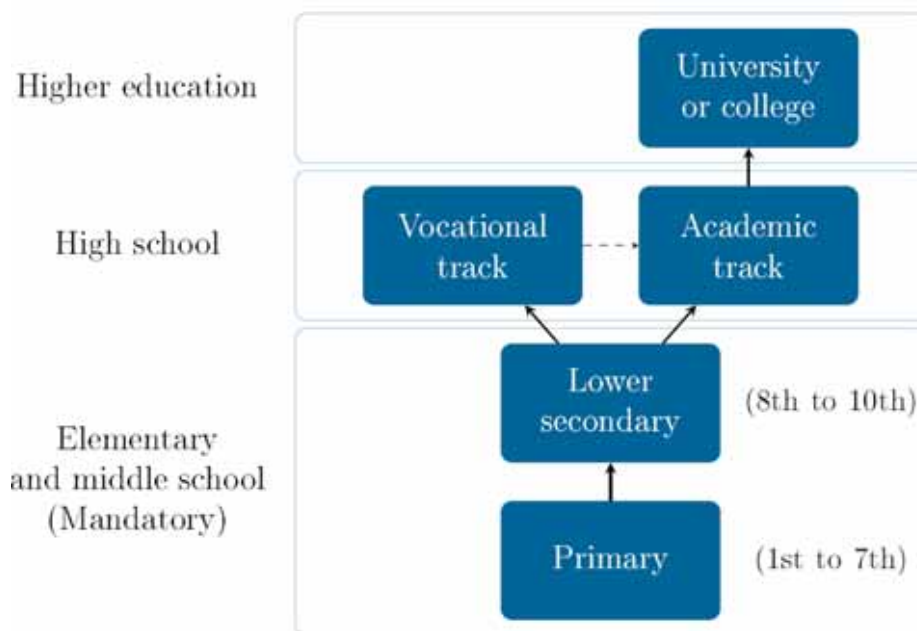


Figure 1 – The Norwegian educational system

Elementary and middle schools are district schools governed by the municipalities, and most children attend a public school in close proximity to their residence.² The curriculum, however, is determined nationally and is similar for all students. Middle schools tend to be larger than elementary schools, implying that children attending different elementary schools feed into the same middle school when starting 8th grade.

Upon graduating from middle school, all children have the right, but no longer a duty, to proceed to an upper secondary education (high school) (the Ministry of Education and Research, 1998). When applying for high school, they can choose between several academic or vocational tracks. The academic tracks consist of three years of schooling, while the vocational tracks consist of three to four years of education and training oriented around specific occupations. The purpose of the academic track is to prepare students for university or college, while the purpose of the vocational track is to prepare students for the labor market. Students graduating from one of the vocational tracks have the option of proceeding to university after attending one additional year of intense academic training (Markussen & Gloppen, 2012), illustrated by the dotted arrow in Figure 1.

High schools are under the responsibility of the county administrations, and the general rule is that students apply for acceptance into a school that offers their desired educational program within their county of residence. If the number of applicants exceeds the number of available slots, the general rule is that acceptance is based on the grade point average (GPA) from middle school.

Students graduating from the academic track in high school have the option of proceeding to university or college (the Ministry of Education and Research, 2005). When applying for higher education, students are

² Only 4 percent of children attend private elementary and middle schools (the Ministry of Education and Research, 1998; Statistics Norway, 2017).

free to apply for acceptance into any university or college they want in the country. As when applying for high school, students apply for acceptance into their desired major. Some majors, however, have special requirements for prior knowledge, meaning that students need to have completed some courses before being accepted (at high school or the equivalent). If no prior knowledge is required or all students applying for the major in that university or college meet the requirements, acceptance is based on the GPA from high school if the number of applicants exceeds the number of available slots.

In the three papers of this thesis, I³ use samples of students at different levels of secondary education (middle school and high school). In the first paper, I investigate the effect of low-achieving peers among students at the start of middle school, while in the second paper we look at the effect of the math teacher's mindset in the last year of middle school. Finally, in the third paper, together with coauthors, I try to change the mindset of students in the first year of high school.

³ Together with coauthors in two of the papers.

3 Conceptual framework

Understanding how to make children thrive in school and exploit more of their potential has engaged scientists for centuries, and my work enters into the field commonly known today as the economics of education. The economics of education is a branch of economics within the more general field of social sciences, which is about understanding human societies and social relationships. The field's history dates back to the eighteenth century, with Adam Smith comparing the education of men to expensive machines in his famous book, commonly referred to as the "*The Wealth of Nations*" (Neal, 2018; Psacharopoulos, 1987).

The economics of education is the study of education from an economic point of view, and focuses especially on the efficiency of educational systems and production of human capital. Children growing up in countries like Norway spend most of their childhood within the educational system, and many remain in the system until well into their twenties. In addition, public funding of the educational sector is immense in many countries. The primary goal of scientists working within the field of the economics of education is to figure out how we can get the optimum return out of these massive investments.

The economics of education field has been developing rapidly ever since the Coleman et al. (1966) report "*Equality of Educational Opportunity*". This report directed the spotlight towards mapping educational inputs and outputs using production functions (Hanushek, 2020; Neal, 2018). We use education production functions to understand how different components related to the educational process produce some outcomes. Their main purpose is to help us identify causal effects in order to understand and learn about the most effective ways to facilitate education.⁴ Common inputs into these functions include, but are not

⁴ More on causality and methods used to uncover causal effects in the empirical approach section below.

limited to, different student, family, peer, teacher and school characteristics, and we usually present them as some version of the following equation:

$$A = f(I, H, P, T, S)$$

where A represents some outcome, usually related to the educational process, like student achievement. I is a vector of student characteristics (e.g. age, gender, "innate ability"), while H represents a vector for family background characteristics (e.g. parental education and earnings, whether the parents are married, access to books and computers at home). P represents a vector for the characteristics of a student's peers (e.g. peers' average achievement or their background characteristics), while T and S represent vectors for teacher (e.g. teacher quality, education, experience) and school (e.g. access to books and computers at school, school facilities) characteristics, respectively. However, not all of these inputs have to be present in order to make it an education production function.

In the three papers of this thesis, I use some version of the education production function to examine different aspects of students' peers and teachers. When investigating the effect of peers, I focus on low-achievers in particular, whereas the focus when looking at teachers is on their mindset and the effect they have on developing the mindsets of their students. Below I describe important aspects of the literatures in this thesis, i.e. the peer effects and the mindset literatures, as well as how the papers in this thesis contribute to moving the literatures one step further.

3.1 Peer effects

In addition to directing the spotlight towards using a production function when mapping educational inputs and outputs, the Coleman et al. (1966) report also highlighted the potential importance of peers. Since then, scientists within the educational literature have investigated several

aspects where they believe students can have an effect on the outcomes of others. The investigated areas include, but are not limited to, test scores, grades, drug use, alcohol consumption, teen pregnancy, dropping out of school and attending college. In a summary of existing literature thus far, Sacerdote (2011) concludes that students' roommates when attending university or college could affect social outcomes, but the effects on GPA are considered modest. Within the literature focusing on test scores in primary and secondary education, however, there does not seem to be any agreement about how students affect each other. Yet, there seems to be consensus that the prevailing model in the literature, the linear-in-means,⁵ might not be the most interesting model when looking at peer effects (Hoxby & Weingarth, 2005; Sacerdote, 2011).

Hoxby & Weingarth (2005) summarize and present some alternative models based on student behavior, including the Bad Apple model of peer effects. The Bad Apple model suggests that the presence of a single student, for instance a low-achiever, has a negative effect on the outcomes of others. Intuitively, we might expect such a student to require more of the teacher's attention compared to other students in his or her class. Consequently, other students might get less help if being exposed to a low-achieving peer in the classroom, which in turn can affect what they learn in school.

Existing literature oriented around the Bad Apple model does not present any clear conclusion about the model. Some studies suggest that including disadvantaged or disruptive students in mainstream classrooms has a negative effect on the performance of others (Carrell et al., 2018; Kristoffersen et al., 2015). Other studies, focused more on inclusive education and special needs, however, find either no effect or small, positive but insignificant effects from sharing a classroom with special education students or disabled students (Hanushek et al., 2002; Ruijs,

⁵ Suggesting that the outcome of a student is a linear function of the mean of his or her peers' outcomes (Sacerdote, 2011).

2017; Ruijs et al., 2010). In the first paper of this thesis, I contribute to the literature on peer effects and investigate the effect of being exposed to at least one low-achieving peer in a new context and approach.

3.2 Mindset

Scientists have, for decades, been concerned with motivation and the role it plays in learning, and have found that people have different beliefs in their own abilities to learn (e.g. Burnette et al., 2013; Dweck, 2006; Dweck & Leggett, 1988; Yeager & Dweck, 2012).⁶ In the mindset literature, we often talk about two kinds of mindset with a spectrum in between. Towards one end of the spectrum, we find people holding more of a growth mindset, who believe that it is possible to develop and change their talents and competencies. People with a growth mindset are process oriented and driven by the desire to develop existing and acquire new knowledge. Towards the other end of the spectrum, we find people with a more fixed mindset. People with a fixed mindset are firm believers in natural talents, and believe that having to put in effort confirms that they do not have what it takes to succeed at a given task. People with a fixed mindset are performance oriented and driven by the desire to continuously prove their talents to themselves and others.

Scientists, primarily within the psychology literature, have looked at how mindsets are developed (Dweck, 2006; Haimovitz & Dweck, 2017; Heggart, 2017; Mueller & Dweck, 1998; Rattan et al., 2015). Among these, there almost seems to be an inherent assumption that teachers can affect the mindset of their students, with papers outlining the importance of employing learning tasks that encourage a growth mindset or that teachers themselves aspire to develop such a mindset (Dweck, 2006; Heggart, 2017). The scientific evidence on the impact, however, seems to be lacking, as there is only limited evidence to date on the impact of teacher mindset on student outcomes (Brooks & Goldstein, 2008). In the

⁶ Definition that follows based on the cited references.

second paper of this thesis, the coauthors and I contribute to this literature by looking at this from an economic point of view. We do this by investigating how teachers' mindset affects outcomes among Norwegian high school students using the education production function.

Another part of the literature, more concerned with changing students' mindsets, has presented evidence that it is possible to move students from a fixed mindset to adapting more of a growth mindset, using interventions designed for this specific purpose (Andersen & Nielsen, 2016; Bettinger et al., 2018; Blackwell et al., 2007; Good et al., 2003; Mueller & Dweck, 1998; Paunesku et al., 2015; Yeager, Romero, et al., 2016; Yeager, Walton, et al., 2016). The interventions that succeed in doing so provide the participants with information about what happens in the brain when you learn, and how using different strategies and perseverance in the light of struggle have helped people move forward. They also relate the content to concepts that the students are already familiar with, like how muscles get stronger when you exercise, and highlight the importance of struggling and making mistakes as a part of the learning process. However, to date, the mindset interventions with positive effects do not include teachers, which could be beneficial for scaling the interventions. Scaling effective interventions is, from an economic point of view, important for human capital development (Heckman, 2000). Moreover, involving the teachers could also help reinforce positive effects of mindset interventions by teachers changing their practice (Kamins & Dweck, 1999; T. Kane et al., 2011; Mueller & Dweck, 1998). Finally, involving the teachers could help legitimize the interventions and thereby promote the growth mindset as something that is important and achievable by the students (Walton & Yeager, in press). In the third paper of the thesis, again together with coauthors, I contribute to this literature by including teachers in an intervention designed to increase the prevalence of a growth mindset among students, and investigate the effect of the intervention using an RCT.

Conceptual framework

4 Empirical approach

The slogan "Correlation does not imply causation" is one of the very first things you learn as an empirical scientist, as uncovering causal effects is at the very core of most empirical research. According to the Oxford Dictionary (2020a), correlation can be defined as a mutual relationship between two or more variables. Causality, on the other hand, can be defined as the relationship between cause and effect (Oxford Dictionary, 2020b).

For centuries, philosophers have argued about and discussed causality and what it entails, and even questioned whether causalities actually exist (R. Kane, 2005). Some argue that all events are predetermined, and consequently there exists no free will, while others argue that everything happens by chance (and of course, everything in between). The lengthy discussion of causality from a philosophical view goes far beyond the focus in this thesis, and I would not be able to do it justice in this brief subsection. Although unable to reach a consensus, among those who believe that causalities actually exist – which is a prerequisite for all the papers in this thesis – there is a general understanding that the cause has to happen before and be sufficient⁷ in leading to the effect (Hoover, 2008; Reiss, 2013).

4.1 Using observational data and experiments to uncover causal effects⁸

A challenge when identifying causation is that we are not able to observe the causal force, and therefore separating causation from mere correlation can be rather difficult. The separation, however, is important

⁷ The Principle of Sufficient Reason.

⁸ The discussion in this section is based on discussions in Angrist & Pischke (2008), Heckman (2005), Hoover (2008), Pearl & Mackenzie (2018) and Reiss (2013), and follow the reasoning in the potential outcomes framework (Rubin, 1974, 1977) closely.

because very often there is no specific reason why events or variables are correlated, they just happen to be, or there might be some common unseen factor through which they are associated. One example is the relationship between ice cream sales and shark attacks (Siegel, 2019). If you plot them in the same graph, it would seem as if ice cream consumption increases shark attacks. However, what you do not observe in the graph is that shark attacks are more likely to happen in the summer when the weather is nice and people go swimming, which is also when consumption of ice cream increases. Identifying this as a correlation, as opposed to a causal effect, is important because otherwise we might end up concluding that the best way to prevent shark attacks is by prohibiting people from buying ice cream.

Uncovering "a causal effect" has several requirements. First, we need to identify one variable to be the "cause" (or the "treatment", often labeled X). Second, we need to identify another variable which is affected by the "cause" (some outcome, often labeled Y). The "effect" we are trying to uncover is the change in outcome (the Y). Third, in order to measure the causal effect, we need to observe a set of possible outcomes (Y), associated with different values of the cause (X). That is, in the case of an experiment, we need to observe the values of Y for someone who has undergone the treatment (where $X = 1$) and for someone who has not (where $X = 0$). Fourth, except the X (the "cause" or the "treatment"), everything else should be equal between the observations if we are to conclude that an effect is causal and not merely correlational.

To illustrate this, assume for the sake of simplicity that X is a binary variable; either someone attends college ($X = 1$) or they do not ($X = 0$); either someone participates in job market training ($X = 1$) or they do not ($X = 0$); either someone is exposed to a low-achieving peer ($X = 1$) or they are not ($X = 0$). What then is the effect of attending college (X) on

future earnings (Y)?⁹ Ideally, to make sure that the "everything else being equal" assumption is met, we would want to observe the outcome for a person (future earnings) both if attending college and if not. This brings us to the fundamental challenge in identifying causal effects; we only observe the outcome (Y) associated with the realized X . We do not, for instance, observe future earnings associated with going to college for someone who starts working after high school. Nor do we observe future earnings of joining the labor force straight after high school for a college graduate. In other words, we only observe the achieved outcome. We only observe the outcome (Y) for either getting the treatment ($X = 1$) or not ($X = 0$), for each individual. The other outcome, the one we do not observe, is the counterfactual.

In the natural sciences, causal effects are typically uncovered by running repeated experiments in the laboratory, changing one variable (X) at a time in order to create and observe something very close to the counterfactual outcome. A famous example of this is the theory of gravity. The benefit of the natural sciences is that the experiments can be re-run as many times as needed, in a strictly controlled environment, easily justifying the "everything else being equal" assumption. You can, for instance, drop an apple multiple times, and if not changing anything about the setting in which it is dropped, the apple will always behave in the same way. Thus, making it possible to change one variable at a time (like the height the apple is dropped from, the size of the apple or the place in which the apple is dropped) and observe the outcome associated with that particular change.

People, however, change behavior and learn from previous experiences,¹⁰ which means it is not possible to do the same within the social sciences. That is, we are not able to change one variable at a time

⁹ Or getting job market training (X) on becoming employed (Y) or being exposed to a low-achieving peer (X) on achievement (Y).

¹⁰ Their own as well as that of others.

in a strictly controlled setting, making the "everything else being equal" assumption more difficult to defend. In order to uncover causal effects, we therefore need to generate a counterfactual outcome that is as realistic as possible. One solution is to compare groups of individuals, with different realized outcomes, and let them represent each other's counterfactuals. That is, by using other individuals as the counterfactuals or as a means to describe potential outcomes. It is common to think of this as a control group and a treatment group.

This, however, poses potential problems. In natural settings, people tend to sort into groups, meaning that individuals in the groups might be different kinds of people, which could potentially bias our estimates (Heckman, 2008). For instance, if the participating subjects in one group have parents with a higher education than the participating subjects in another group, the result from the comparison might be partly because of parental education. That is, the results might pick up something that is inherently different between the two groups with regards to the education of their parents, not just the effect of the "cause" or the "treatment".

An example of this can be illustrated by investigating the effect of attending private schools on student outcomes. A simple comparison of test scores of children attending private schools to children attending public schools is likely to yield biased estimates. The reason is that educated parents might value education more, and as a result may help their children more with their school work and also be more likely to send their children to private schools. Thereby, introducing a confounder – parental education – that could potentially affect both the behavior of the participants in general (helping their children more with school work) and their choice to participate in the "treatment" (sending their children to private schools), leading to the problem of omitted variable bias or selection bias in the estimated effects due to unobserved differences between the groups. Consequently, the two groups represent a poor counterfactual to each other.

In order to uncover the true causal effect of the "treatment", it is important to control for these potential confounders so that we are actually comparing apples and apples, not apples and oranges. If we imagine that the scientist was able to observe all possible background characteristics, it would be feasible to control for these differences. In the case mentioned above, with the confounder of parental education, this is fairly easy; the scientist only needs to observe the education of the participating subjects' parents. It can, however, be tricky, since it is not always possible to observe such confounders. Examples of this include the effect of ability or social and emotional skills on school or labor market outcomes. One can observe possible proxies for ability or social and emotional skills, but never a person's entire ability or the entire range of social and emotional skills, as these are way too complex to be measured in full. Therefore, we would never be able to fully control for these aspects when looking at school or labor market outcomes. As a result, the groups could be inherently different with regards to unobservable baseline characteristics.¹¹

One way of ensuring that the groups are similar (on baseline characteristics), and consequently good representations of each other's counterfactuals, is to introduce random variation in who receives the "treatment". The idea behind a randomized design is that there should not be, on average, any other differences between the groups than the actual treatment (Cartwright, 2012). Randomization is meant to deal with this by creating groups so that all other factors that might influence outcomes are identically distributed among the groups. An important element of randomization is that it is not the participating subjects who decide whether to be offered the treatment, but some exogenous factor.

This can be done using a randomized controlled trial (RCT). RCTs are by many considered to be the perfect statistical tool for uncovering causality, and are often referred to as the scientific gold standard in the

¹¹ Other such characteristics include, but are not limited to, genetics, environment, etc.

evidence-based areas of the literature.¹² An RCT is a specific kind of controlled experiment aimed at identifying the causal effect of some treatment, where the scientist in charge of the experiment has full control over the randomization, that is, who receives the treatment. When the participating subjects are randomly assigned to control and treatment, the two groups will be similar. This can be done using methods such as a computerized random number generator, a lottery, etc. After randomization, participating subjects assigned to the treatment group are exposed to some treatment, while the subjects in the control group receive either an alternative treatment, a placebo treatment or no treatment at all. Afterwards, the observed outcomes of participating subjects in the control and treatment group are compared, and the difference between them is considered the causal effect of the treatment.

However, there are situations where it is not possible to use RCTs to identify causal effects because of moral or ethical considerations. When dealing with potential destructive or fatal outcomes for instance, like the effect of smoking on cancer, drinking when driving, using illegal substances on health outcomes, etc., it is not possible to defend assigning some participants to the treatment group from a moral or ethical perspective as the treatment is potentially harmful. In other cases, conducting an RCT could be too costly, as uncovering treatment effects might require large samples.

A different approach to ensuring similar groups and random variation in who receives the "treatment" is to look for situations with some kind of exogenous shock that mimic a randomized experiment. These situations might arise if human institutions or natural forces create something that is close to random assignment (Angrist & Krueger, 1999). The main difference from an RCT is that the scientist is no longer in charge of who receives the "treatment" and who is placed in the control group.

¹² However, despite having the reputation of being the gold standard, there are a lot of pitfalls one can encounter when doing an RCT. This is to some extent illustrated in the third paper of this thesis.

However, the important element still being that it is not the participating subjects who choose whether to get access to the "treatment" or not. An example of this is the use of hurricanes in the US to investigate the structure of peer effects (Imberman et al., 2012). Because of the damage caused by two hurricanes, over one million people were forced to evacuate. The school districts surrounding the affected areas enrolled children as fast as possible. Evacuated children came from some of the lowest-performing schools in the country, and as a result, some cities received large numbers of low-performing students. This created concerns among the non-evacuee population that their children would be negatively affected by the inflow of low-performing students. However, there was nothing they could do to stop it from happening,¹³ they just had to sit back and receive the "treatment".

Another example is reflected in a distinctive feature of the Norwegian educational system; the random assignment of students to classes in elementary and middle schools, which is used to emulate an experiment in the first and second paper in this thesis. When starting school in Norway, the law states that students should be assigned to classes based on their need for social belonging, and that they should not be placed in groups based on ability, gender or ethnicity (the Ministry of Education and Research, 1998). Guidelines define social belonging as that assignment to classes within a school should reflect the student population,¹⁴ and states that the schools have the final say when it comes to allocation (the Ministry of Education and Research, 2017). This then eliminates the problem of certain parents choosing a particular teacher or group of students, and hence which class their child is assigned to.¹⁵

¹³ At least not in the short term. In the longer term, it would be possible to move.

¹⁴ Meaning that if 30 percent of the students are boys or immigrants, each of the classes should reflect this.

¹⁵ For a more detailed description of the random assignment within schools, see the section about education in Norway in paper one or two.

In the three papers of this thesis, I¹⁶ look for settings that offer a source of exogenous variation in the treatment assignment that is close to random assignment (paper I and II) or randomly assigned students myself (paper III). Subsequently, I use some version of the education production function to identify factors that are relevant in getting children to reach their full potential. In the first paper, I exploit the random assignment to classes in Norwegian elementary and middle schools to investigate the causal effect of being exposed to at least one low-achieving peer in the classroom on students' school achievement. In the second paper, I utilize the same random assignment as in the first paper, and look at how teacher mindset affects student outcomes. In the third paper, I randomly assign students to a control group or a treatment group using an RCT design and investigate if it is possible to change student outcomes using a mindset intervention with specific teacher involvement.

¹⁶ Together with coauthors in two of the papers.

5 Summary of papers

"The effect of low-achieving peers"

Inclusive education, or educating all children in the same classrooms regardless of individual needs or difficulties, has received increasing attention in recent years (Ainscow and César, 2006; Ruijs, 2017; Ruijs et al., 2010). As a result, several countries have moved away from special needs education towards more inclusive environments, where the focus is on educating all students in mainstream classrooms (European Agency for Development in Special Needs Education, 2013; Ogden, 2014). Norway is not an exception, but is in fact considered by many to be a frontrunner when it comes to inclusive education.

Existing literature reports evidence that including disadvantaged or disruptive students in mainstream classrooms might have a negative effect on their peers (Carrell et al., 2018; Carrell and Hoekstra, 2010; Kristoffersen et al., 2015). Inspired by this, and especially the Bad Apple model of peer effects (Hoxby & Weingarth, 2005; Sacerdote, 2011), I examine the effect on achievement from exposure to at least one low-achieving peer in Norwegian middle schools. If including such peers in the mainstream classroom is in fact harmful for students, we should expect those who are exposed to perform worse than those who are not. To investigate this, I utilize the change in class composition when starting middle school and the random assignment of students to classes, and investigate this in an entire cohort of Norwegian students.

I employ comprehensive registry data for the entire Norwegian population, and estimate a fixed effect model where I exploit the differences in students' exposure to low-achieving peers. My analyses suggest that the presence of at least one low-achieving peer in the classroom has no effect on the achievement of other students. The results are consistent when dividing the sample based on subject, gender,

parental education and earnings as well as students' own position on the achievement distribution.

"Girls in STEM: the role of teacher mindset" (with Eric Bettinger, Mari Rege and Ingeborg F. Solli)

While outnumbering men at most college campuses around the world, women remain underrepresented in core science, technology, engineering and math, commonly referred to as the STEM fields (Kahn & Ginther, 2017). The gaps in female representation start early in students' careers, and gender disparities in levels of engagement can already be seen in math as early as age six (Petersen & Hyde, 2017). Despite the difference in representation, several studies report only small differences in average achievement in the subjects most important for continuing towards a STEM degree in college (Hyde et al., 2008; Lindberg et al., 2010; Petersen & Hyde, 2014).

In the second paper of this thesis, we investigate whether the mindset of middle school math teachers plays a role when it comes to students' engagement and achievement in high school math. In addition, we examine the effect on representation in the most challenging math course in high school, a choice highlighted by many as a critical gateway into a STEM career later on (Ma & Johnson, 2008; Sells, 1980). There are also some suggestions in the literature that peer and teacher support are more important for girls and that mindset might be a contributing factor in course selection, especially among girls (Beilock et al., 2010; Blackwell et al., 2007; Good et al., 2012; Nix et al., 2015; Perez-Felkner et al., 2017; Riegler-Crumb et al., 2006; Samuelsson & Samuelsson, 2016). Therefore, we divide the sample based on gender, and have a special focus on girls in this study.

We utilize a unique dataset with extensive survey data for a full cohort of students starting high school in the school year 2017/2018 in two of the largest counties in Norway. The survey data is matched with registry

data from the counties and Statistics Norway. We measure teachers' mindset using students' perceptions of their middle school math teacher's mindset. We look at the effect on each individual student using their peers' perceptions of their math teacher's mindset in a fixed effect approach. Our result suggests that being exposed to a math teacher with more of a growth mindset during the last year of middle school has a positive effect on students' own mindset, as well as attitudes and expectations in relation to math. We also find that exposed students are more likely to choose the more challenging math course in high school. We do not, however, find any effect on students' achievement or that there are different effects based on the gender of the student; being exposed to a teacher with more of a growth mindset is equally beneficial for boys and girls.

"Involving the teachers: effects of a growth mindset intervention with teacher involvement" (with Mari Rege and David S. Yeager)

Believing that intelligence and talents can be developed and changed through effort, perseverance as well as hard work and by using different strategies, is what psychologists refer to as holding a growth mindset (Dweck, 2006; Mueller & Dweck, 1998; Yeager & Dweck, 2012). Several studies report that teaching students about this and the brain's malleability can help them develop such a mindset and thereby improve motivation and outcomes at school (e.g. Aronson et al., 2002; Bettinger et al., 2018; Blackwell et al., 2007; Good et al., 2003; Paunesku et al., 2015; Yeager et al., 2014; Yeager, Romero, et al., 2016). However, none of the interventions reporting positive effects have included teachers as yet.

In the third paper of this thesis, my coauthors and I address this gap in the literature by testing an intervention with specific teacher involvement. We do this by giving teachers information about growth

mindset research, and tools and guidance to help them pass the information on to their students. We expect that in turn, this will make students more likely to choose challenging assignments, show more interest in school activities, have higher expectations in relation to their own performance and increase their achievement and perseverance in school.

We test this using a randomized controlled trial among a self-selected group of classes in several high schools in two counties in Norway. First, we invite teachers to sign up for the experiment before randomly assigning classes, and thus teachers, to a control group and a treatment group. After randomization, we invite teachers in the treatment group to a full day seminar. The seminar includes lectures about the growth mindset and teachers are given an introduction to the tools developed to help them create a classroom climate that supports a growth mindset. The purpose of the full day seminar was to make the teachers confident in conducting an educational program with their students, which they did in the months following the seminar.

In our empirical analyses, we utilize answers to survey questions as well as data from county administrative records, and investigate the effect of the treatment in a specification with educational program and school fixed effects. We find that treated students show more of a growth mindset in general and when it comes to math in particular, compared to students in the control group. We also find evidence suggesting that treated students have a more positive attitude towards challenge, with the most important outcome being the choice of math course in high school. Here we find that among students who are able to choose which math course to participate in, treated students are more likely to choose the more advanced course than their counterpart in the control group.

References

Ainscow, M., César, M. (2006). Inclusive education ten years after Salamanca: Setting the agenda. *European Journal of Psychology of Education, 21* (3), 231–238.

Andersen, S. C., & Nielsen, H. S. (2016). Reading intervention with a growth mindset approach improves children's skills. *Proceedings of the National Academy of Sciences, 113* (43), 12111-12113.

Angrist, J. D., & Krueger, A. B. (1999). Empirical Strategies in Labor Economics. In *Handbook of Labor Economics* (Vol. 3, pp. 1277-1366). The Netherlands: North-Holland, an imprint of Elsevier.

Angrist, J. D., & Pischke, J. S. (2008). *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton: Princeton University Press.

Aronson, J., Fried, C. B., & Good, C. (2002). Reducing the Effects of Stereotype Threat on African American College Students by Shaping Theories of Intelligence. *Journal of Experimental Social Psychology, 38* (2), 113-125.

Beilock, S. L., Gunderson, E. A., Ramirez, G., & Levine, S. C. (2010). Female teachers' math anxiety affects girls' math achievement. *Proceedings of the National Academy of Sciences, 107* (5), 1860-1863.

Bettinger, E. P., Ludvigsen, S., Rege, M., Solli, I. F., & Yeager, D. (2018). Increasing perseverance in math: Evidence from a field experiment in Norway. *Journal of Economic Behavior & Organization, 146*, 1-13.

Blackwell, L. S., Trzesniewski, K. H., & Dweck, C. S. (2007). Implicit Theories of Intelligence Predict Achievement Across an Adolescent Transition: A Longitudinal Study and an Intervention. *Child Development, 78* (1), 246-263.

References

- Brooks, R., & Goldstein, S. (2008). The Mindset of Teachers Capable of Fostering Resilience in Students. *Canadian Journal of School Psychology, 23* (1), 114-126.
- Burnette, J. L., O'Boyle, E. H., VanEpps, E. M., Pollack, J. M., & Finkel, E. J. (2013). Mind-sets matter: A meta-analytic review of implicit theories and self-regulation. *Psychological Bulletin, 139* (3), 633-701.
- Carrell, S. E., Hoekstra, M., & Kuka, E. (2018). The long-run effects of disruptive peers. *American Economic Review, 108* (11), 3377-3413.
- Carrell, S.E., 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.
- Carrell, S. E., Page, M. E., & Vest, J. E. (2010). Sex and Science: How Professor Gender Perpetuates the Gender Gap. *The Quarterly Journal of Economics, 125* (3), 1101-1144.
- Cartwright, N. (2012). *RCTs, evidence, and predicting policy effectiveness*. Oxford: Oxford University Press.
- Chetty, R., Friedman, J. N., Hilger, N., Saez, E., Schanzenbach, D. W., & Yagan, D. (2011). How Does your Kindergarten Classroom Affect your Earnings? Evidence from Project STAR. *Quarterly Journal of Economics, 126* (4), 1593-1660.
- Coleman, J. S., Campbell, E. Q., Hobson, C. J., McPartland, J., Mood, A. M., Weinfeld, F. D., & York, R. L. (1966). *Equality of Educational Opportunity*. Washington, D.C.: U.S. Government Printing Office.
- Deming, D. J. (2017). The Growing Importance of Social Skills in the Labor Market. *The Quarterly Journal of Economics, 132* (4), 1393-1640.
- Dweck, C. S. (2006). *Mindset: The New Psychology of Success*. New York: The Random House Publishing Group.

References

- Dweck, C. S. (2009). Who Will the 21st-Century Learners Be? *Knowledge Quest*, 38 (2), 8-9.
- Dweck, C. S., & Leggett, E. L. (1988). A social-cognitive approach to motivation and personality. *Psychological Review*, 95 (2), 236-273.
- Edin, P. A., Fredriksson, P., Nybom, M., & Ockert, B. (2017). The Rising Return to Non- cognitive Skill. *IZA Discussion Paper No. 10914*.
- European Agency for Development in Special Needs Education (2013). *Organisation of provision to support inclusive education. Literature review*. Odense, Denmark. Retrieved from https://www.european-agency.org/sites/default/files/organisation-of-provision-to-support-inclusive-education-2013-literature-review_Organisation-of-Provision-Literature-Review.pdf
- Freeman, P., & Aspray, V. (1999). *The Supply of Information Technology Workers in the United States* (Tech. Rep.). ERIC.
- Good, C., Aronson, J., & Inzlicht, M. (2003). Improving adolescents' standardized test performance: An intervention to reduce the effects of stereotype threat. *Journal of Applied Developmental Psychology*, 24 (6), 643-662.
- Good, C., Rattan, A., & Dweck, C. S. (2012). Why do women opt out? Sense of belonging and women's representation in mathematics. *Journal of Personality and Social Psychology*, 102 (4), 700-717.
- Haimovitz, K., & Dweck, C. S. (2017). The Origins of Children's Growth and Fixed Mindsets: New Research and a New Proposal. *Child Development*, 88 (6), 1849-1839.
- Hamre, B. K., & Pianta, R. C. (2005). Can Instructional and Emotional Support in the First-grade Classroom Make a Difference for Children at Risk of School Failure? *Child Development*, 76 (5), 949-967.

References

- Hanushek, E. A. (2020). Education production functions. In *The Economics of Education* (Second Edition) (pp. 161-170). US/UK: Academic Press, an imprint of Elsevier.
- Hanushek, E. A., Kain, J. F., & Rivkin, S. G. (2002). Inferring Program Effects for Special Populations: Does Special Education Raise Achievement for Students with Disabilities? *Review of Economics and Statistics*, 84 (4), 384-399.
- Heckman, J. J. (2000). Policies to foster human capital. *Research in Economics*, 54 (1), 3-56.
- Heckman, J. J. (2005). The Scientific Model of Causality. *Sociological Methodology*, 35 (1), 1-97.
- Heckman, J. J. (2008). Econometric Causality. *International Statistical Review*, 76 (1), 1-27.
- Heckman, J. J., & Kautz, T. (2012). Hard evidence on soft skills. *Labour Economics*, 19 (4), 431-464.
- Heckman, J. J., Stixrud, J., & Urzua, S. (2006). The Effects of Cognitive and Noncognitive Abilities on Labor Market Outcomes and Social Behavior. *Journal of Labor Economics*, 24 (3), 411-482.
- Heggart, K. (2017). Early childhood student teachers: It pays to belong to the IEU. *Newsmonth*, 37 (7).
- Hoover, D. K. (2008). Causality in Economics and Econometrics. In *Palgrave Macmillan (eds) The New Palgrave Dictionary of Economics*. London: Palgrave Macmillan.
- Howes, C., Hamilton, C. E., & Matheson, C. C. (1994). Children's Relationships with Peers: Differential Associations with Aspects of the Teacher-Child Relationship. *Child Development*, 65 (1), 253-263.

References

- Hoxby, C. M., & Weingarth, G. (2005). Taking race out of the equation: School reassignment and the structure of peer effects. *Unpublished manuscript*.
- Hyde, J. S., Lindberg, S. M., Linn, M. C., Ellis, A. B., & Williams, C. C. (2008). Gender Similarities Characterize Math Performance. *Science*, *321* (5888), 494-495.
- Imberman, S. A., Kugler, A. D., & Sacerdote, B. I. (2012). Katrina's Children: Evidence on the Structure of Peer Effects from Hurricane Evacuees. *American Economic Review*, *102* (3), 2048-2082.
- Kahn, S., & Ginther, D. (2017). Women and STEM. *NBER Working Paper No. 23525*.
- Kamins, M. L., & Dweck, C. S. (1999). Person versus process praise and criticism: Implications for contingent self-worth and coping. *Developmental Psychology*, *35* (3), 833-847.
- Kane, R. (2005). *A contemporary introduction to free will*. Oxford: Oxford University Press.
- Kane, T. J., Taylor, E. S., Tyler, J. H., & Wooten, A. L. (2011). Identifying Effective Classroom Practices Using Student Achievement Data. *Journal of Human Resources*, *46* (3), 387-613.
- Kautz, T., Heckman, J. J., Diris, R., Ter Weel, B., & Borghans, L. (2014). Fostering and Measuring Skills: Improving Cognitive and Non-Cognitive Skills to Promote Lifetime Success. *NBER Working Paper No. 20749*.
- Kristoffersen, J. H. G., Krægpøth, M. V., Nielsen, H. S., & Simonsen, M. (2015). Disruptive school peers and student outcomes. *Economics of Education Review*, *45*, 1-13.

References

Lindberg, S. M., Hyde, J. S., Petersen, J. L., & Linn, M. C. (2010). New trends in gender and mathematics performance: A meta-analysis. *Psychological Bulletin*, 136 (6), 1123-1135.

Ma, X., & Johnson, V. (2008). Mathematics as the critical filter: Curricular effects on gendered career choices. In H. M. G. Watt & J. S. Eccles (Eds.), *Gender and occupational outcomes: Longitudinal assessments of individual, social, and cultural influences* (pp. 33-83). Washington, DC: American Psychological Association.

Markussen, E., & Gloppen, S. K. (2012). *Påbygg – et gode eller en nødløsning? En studie av påbygging til generell studiekompetanse i Østfold, Akershus, Buskerud, Rogaland og Nord-Trøndelag skoleåret 2010-2011* (Tech. Rep.). Retrieved from <https://nifu.brage.unit.no/nifu-xmlui/bitstream/handle/11250/280848/NIFUrapport2012-2.pdf?sequence=1&isAllowed=y>

Ministry of Education and Research. (1998). *Lov 17. juli 1998 om grunnskolen og den vidaregåande opplæringa (the Education Act)*. Retrieved from <https://lovdata.no/dokument/NL/lov/1998-07-17-61> (Opplæringslova)

Ministry of Education and Research. (2005). *Lov 13. desember 2005 om universiteter og høyskoler (Act relating to universities and university colleges)*. Retrieved from <https://lovdata.no/dokument/NL/lov/2005-04-01-15> (Universitets- og høyskoleloven)

Ministry of Education and Research. (2017). *Veiledning om organisering av elevene*. Online. Retrieved from www.regjeringen.no/contentassets/f94154aa3d2b491ba1ac2f7f658cb019/veiledning-om-organisering-av-elevne_oppdatert-april-2017.pdf (Accessed November 2018)

References

- Mueller, C. M., & Dweck, C. S. (1998). Praise for Intelligence Can Undermine Children's Motivation and Performance. *Journal of Personality and Social Psychology*, 75 (1), 33-32.
- Neal, D. A. (2018). *Information, Incentives, and Education Policy*. Cambridge: Harvard University Press.
- Nix, S., Perez-Felkner, L., & Thomas, K. (2015). Perceived mathematical ability under challenge: a longitudinal perspective on sex segregation among STEM degree fields. *Frontiers in Psychology*, 6, Article 530.
- OECD Publishing. (2015). *Skills for Social Progress: The Power of Social and Emotional Skills*. Paris: OECD Publishing. (OECD Skills Studies)
- Ogden, T. (2014). Special needs education in Norway – the past, present, and future of the field. *Special Education Past, Present, and Future: Perspectives from the Field* (Advances in Learning and Behavioral Disabilities, Vol. 27), Emerald Group Publishing Limited, pp. 213-238.
- Oxford Dictionary. (2020a). *Correlation*. Retrieved from <https://www.lexico.com/en/definition/correlation> (Accessed March 2020)
- Oxford Dictionary. (2020b). *Causality*. Retrieved from <https://www.lexico.com/en/definition/causality> (Accessed March 2020)
- Paunesku, D., Walton, G. M., Romero, C., Smith, E. N., Yeager, D. S., & Dweck, C. S. (2015). Mind-Set Interventions Are a Scalable Treatment for Academic Underachievement. *Psychological Science*, 26 (6), 784-793.
- Pearl, J., & Mackenzie, D. (2018). *The Book of Why: The New Science of Cause and Effect*. UK: Penguin Books.

References

- Perez-Felkner, L., Nix, S., & Thomas, K. (2017). Gendered Pathways: How Mathematics Ability Beliefs Shape Secondary and Postsecondary Course and Degree Field Choices. *Frontiers in Psychology, 8*, Article 386.
- Petersen, J. L., & Hyde, J. S. (2014). Gender-Related Academic and Occupational Interests and Goals. In *Advances in Child Development and Behavior* (Vol. 47, pp. 43-76). US/UK: Academic Press, an imprint of Elsevier.
- Petersen, J. L., & Hyde, J. S. (2017). Trajectories of self-perceived math ability, utility value and interest across middle school as predictors of high school math performance. *Educational Psychology, 37* (4), 438-456.
- Psacharopoulos, G. (1987). *Economics of Education: Research and Studies*. UK: Pergamon Press.
- Rattan, A., Savani, K., Chugh, D., & Dweck, C. S. (2015). Leveraging Mindsets to Promote Academic Achievement: Policy Recommendations. *Perspectives on Psychological Science, 10* (6), 721-726.
- Reiss, J. (2013). *Philosophy of Economics: A Contemporary Introduction*. US: Routledge.
- Riegle-Crumb, C., Farkas, G., & Muller, C. (2006). The Role of Gender and Friendship in Advanced Course Taking. *Sociology of Education, 79* (3), 206-228.
- Rockoff, J. E. (2004). The Impact of Individual Teachers on Student Achievement: Evidence from Panel Data. *American Economic Review, 94* (2), 247-252.

References

- Rubin, Donald B. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology*, 66 (5), 688-701.
- Rubin, Donald B. (1977). Assignment to treatment group on the basis of a covariate. *Journal of Educational Statistics*, 2 (1), 1-26.
- Ruijs, N. M. (2017). The impact of special needs students on classmate performance. *Economics of Education Review*, 58, 13-31.
- Ruijs, N. M., Van der Veen, I., & Peetsma, T. T. (2010). Inclusive education and students without special educational needs. *Educational Research*, 52 (4), 331-390.
- Sacerdote, B. (2011). Peer Effects in Education: How Might They Work, How Big Are They and How Much Do We Know Thus Far? In *Handbook of the Economics of Education* (Vol. 3, pp. 249-277). The Netherlands: North-Holland, an imprint of Elsevier.
- Samuelsson, M., & Samuelsson, J. (2016). Gender differences in boys' and girls' perception of teaching and learning mathematics. *Open Review of Educational Research*, 3 (1), 18-34.
- Sells, L. V. (1980). Mathematics: The Invisible Filter. *Engineering Education*, 70 (4), 340-341.
- Siegel, E. (2019). *Why Ice Cream Is Linked to Shark Attacks – Correlation/Causation – The Dr. Data Show*. Retrieved from <https://www.youtube.com/watch?v=AOB6vcwFL-I> (Accessed March 2020)
- Statistics Norway. (2017). *Elevar i grunnskolen*. Retrieved from <https://www.ssb.no/utdanning/statistikker/utgrs> (Table 4 accessed November 2018)

References

Walton, G. M., & Yeager, D. S. (in press). Seed and Soil: Psychological Affordances in Contexts Help to Explain Where Wise Interventions Succeed or Fail. *Current Directions in Psychological Science*.

Yeager, D. S., & Dweck, C. S. (2012). Mindsets That Promote Resilience: When Students Believe That Personal Characteristics Can Be Developed. *Educational Psychologist*, 47 (4), 302-314.

Yeager, D. S., Hanselman, P., Walton, G. M., Murray, J. S., Crosnoe, R., Muller, C., ... others (2019). A national experiment reveals where a growth mindset improves achievement. *Nature*, 573 (7774), 364-369.

Yeager, D. S., Johnson, R., Spitzer, B. J., Trzesniewski, K. H., Powers, J., & Dweck, C. S. (2014). The far-reaching effects of believing people can change: Implicit theories of personality shape stress, health, and achievement during adolescence. *Journal of personality and social psychology*, 106 (6), 867-884.

Yeager, D. S., Romero, C., Paunesku, D., Hulleman, C. S., Schneider, B., Hinojosa, C., ... others (2016). Using design thinking to improve psychological interventions: The case of the growth mindset during the transition to high school. *Journal of Educational Psychology*, 108 (3), 374-391.

Yeager, D. S., Walton, G. M., Brady, S. T., Akcinar, E. N., Paunesku, D., Keane, L., ... others (2016). Teaching a lay theory before college narrows achievement gaps at scale. *Proceedings of the National Academy of Sciences*, 113 (24), 3341-3348.

Paper I - III

Paper I: Auestad, M. L. (2018). The effect of low-achieving peers. *Labour Economics*, 55, 178-214.

Paper II: Girls in STEM: the role of teacher mindset (with Eric Bettinger, Mari Rege and Ingeborg F. Solli).

Paper III: Involving the teachers: effects of a growth mindset intervention with teacher involvement (with Mari Rege and David S. Yeager).

Paper I





Contents lists available at ScienceDirect

Labour Economics

journal homepage: www.elsevier.com/locate/labecoThe effect of low-achieving peers[☆]

May Linn Auestad*

University of Stavanger, Stavanger, 4021, Norway



ARTICLE INFO

JEL classification:

I21
J13

Keywords:

Peer effects
Low-achievers
Special educational needs
Inclusive education

ABSTRACT

Using Norwegian registry data, I examine the effect that low-achieving pupils have on their peers in the inclusive Norwegian compulsory education. I exploit the fact that pupils in Norway have to change school when starting lower secondary, and that national tests are conducted in the same subjects, both before and after this change. I focus on the exposure to at least one low-achieving pupil in a fixed effect approach. My results suggest that low-achieving pupils do not have any effect on their peers, and are even consistent when dividing pupils into different groups based on subject, gender, socioeconomic background and pupils' own position in the achievement distribution.

1. Introduction

Inclusive education, or educating all children in the same classrooms regardless of individual needs or difficulties, has received increasing attention in recent years (Ainscow and César, 2006; Ruijs, 2017; Ruijs and Peetsma, 2009; Ruijs et al., 2010). Several countries have moved away from special needs schools towards a more inclusive environment within the compulsory education, and the focus is on educating all pupils in mainstream classrooms (European Agency for Development in Special Needs Education, 2013; Ogden, 2014). In 1994, more than 300 participants from 92 governments and 25 international organisations met in Salamanca, Spain in order to promote the approach of inclusive education, and adopted “the Salamanca Statement and Framework for Action on Special Needs Education” (Ministry of Education and Science and United Nations Educational, Scientific and Cultural Organization, 1994), which encourages governments to give the highest priority to inclusive education. Existing literature suggests that including disadvantaged and disruptive peers in mainstream classrooms has a negative effect on the performance of other pupils (Carrell et al., 2016; Carrell and Hoekstra, 2010; Figlio, 2007; Kristoffersen et al., 2015). Other studies find small, positive but insignificant or no effects of sharing a classroom with a disabled or special educational needs (SEN) pupil (Hanushek et al., 2002; Ruijs, 2017; Ruijs et al., 2010). My paper adds to this literature by investigating the effect of low-achieving peers in a highly inclusive environment.

Inspired by the Bad Apple model, suggesting that the presence of a single low-achieving pupil has a negative effect on the outcomes for many other pupils,¹ this paper is concerned with the effect of including pupils in the lower tail of the achievement distribution in mainstream classrooms. On the one hand, there are a lot of potential problems that might arise in an inclusive education setting, e.g. low-achieving pupils might require more of the teacher's attention, possibly at the expense of other pupils (Lavy et al., 2012a; Ruijs, 2017; Ruijs and Peetsma, 2009; Ruijs et al., 2010). The general level of education might be lowered so that all pupils are able to follow, making other pupils worse off than they would have been if they were in a different class setting (Ruijs and Peetsma, 2009). Low-achieving pupils may be more disruptive than higher achieving pupils, hence increasing the level of disruption and maybe even violence in the classroom, and thereby reducing the time spent teaching or the time pupils spend working on their own (Aizer, 2008; Carrell et al., 2016; Carrell and Hoekstra, 2010; Figlio, 2007; Lavy et al., 2012a; Lavy and Schlosser, 2011; Lazear, 2001), or other pupils might copy undesirable traits and behaviours from their low-achieving peers (Ruijs et al., 2010). On the other hand, there might be positive effects from inclusive environments because the education could be more specialised towards the individual pupil, extra teacher resources could benefit all pupils, and pupils may become less dependent on the teacher if they receive less help and learn to study independently (Ruijs and Peetsma, 2009; Ruijs et al., 2010). In addition, it could be reasoned that inclusive education helps avoid discrimination towards low-achieving

[☆] I am grateful to the Editor, two anonymous reviewers, Mari Rege, Aline Buetikofer, Marianne P. Bitler, Marianne E. Page, Eric Bettinger, Scott E. Carrell, Paco Martorell, Ingeborg F. Solli, Venke F. Haaland, participants at the applied micro brown bag at UC Davis and seminar participants at University of Stavanger for their help and useful comments during the work on this project.

* Tel.: +4793249910.

E-mail address: may.l.auestad@uis.no

<https://doi.org/10.1016/j.labeco.2018.10.004>

Received 8 February 2018; Received in revised form 12 July 2018; Accepted 12 October 2018

Available online 13 October 2018

0927-5371/© 2018 The Author. Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license.

<http://creativecommons.org/licenses/by-nc-nd/4.0/>

pupils because they are included in the classroom, and thereby other pupils get more accustomed to having them around (Ruijs et al., 2010).

I investigate the effect that the presence of at least one low-achieving pupil in the classroom has on other pupils' achievement in the Norwegian compulsory education. I do this by looking at the effect on the test scores in the national tests in lower secondary school 9th grade, using a fixed effect approach based on the education production function. As a source of exogenous variation, I use the fact that pupils in the Norwegian compulsory education have to change schools and are therefore assigned to a new class and exposed to new peers when starting lower secondary, and that they are tested in the same subjects both before and after this change. In particular, I ask whether pupils who have at least one low-achieving peer in their class show different academic results compared to pupils who are not exposed to any low-achievers in lower secondary school, given that they attend the same school in lower secondary and that they attended the same class in primary school. When defining low-achieving pupils, I use test scores from national tests before pupils change schools, in primary school 5th grade, and define low-achievers as pupils who score below the 10th percentile in the national subject- and cohort-specific achievement distribution in at least two of the three core subjects, which is similar to the approach used in Lavy et al. (2012b).

In addition, the registry data I use has enough information to link pupils to their parents, allowing me to control for individual and parental characteristics. The identifying assumptions are that there is variation in being exposed to a low-achieving pupil within the school and that the likelihood of exposure to at least one low-achiever in the classroom is independent of all other factors that may influence school outcomes. The data also contains group identifiers, allowing me to identify peer effects at the classroom level, which previous studies have shown tend to generate much stronger peer effect measures than at the grade level (Burke and Sass, 2013). Many studies in the literature, however, have found this difficult to do in practice because of data constraints (Kristoffersen et al., 2015).

Results show that the presence of at least one low-achieving pupil in the classroom has no effect on other pupils. The results are consistent when dividing the sample into subsamples based on subject, gender, parental education and earnings, and pupils' own position in the achievement distribution.

In a world where inclusion is being implemented in several countries, it is important to understand the effects of including different groups of pupils (like low-achievers). Thus, investigating the effects in Norway, which is considered a frontrunner when it comes to inclusive education, seems highly appropriate. This paper also adds to the literature by addressing some of the concerns raised in the existing literature with regards to the fact that the majority of studies only report mean estimates and the fact that most studies are conducted using data from USA (Kalamouka et al., 2007; Ruijs and Peetsma, 2009). One advantage of my study is that I am able to divide pupils into groups based on characteristics determined prior to the treatment, and thereby investigate the effects on different groups of pupils. In addition, I use a different treatment variable than previous studies in the inclusive education literature, and exploit the fact that low-achievers who do not receive special education may influence their peers to an even greater extent than SEN pupils, because when trying to get their educational needs met, they might require more of the resources allocated to the pupils without special educational needs than the average pupil. In addition, the dataset I use in this study contains information about all pupils in Norway, allowing me to investigate the effects for an entire cohort of Norwegian pupils.

The rest of the paper is organised as follows: In section 2 I describe the institutional settings with regards to the Norwegian compulsory edu-

cation, while section 3 describes the data used in this study. In section 4, I present the empirical approach used, while the empirical results follow in section 5. In section 6, I interpret and discuss the results before finally concluding in section 7.

2. Institutional settings

2.1. Education in Norway

Compulsory education in Norway, lasts for ten years and is divided into primary (1st to 7th grade) and lower secondary (8th to 10th grade) education. Children start primary school the year they turn six and finish the year they turn thirteen, while lower secondary education starts at the age of thirteen and lasts until the age of sixteen. After finishing compulsory schooling, pupils have the opportunity to move up to upper secondary school, where they can choose between several academic or vocational tracks or they can choose to drop out of school. The vocational tracks last for three to four years and are intended to help pupils learn different skills (based on the chosen occupation) in order to start working afterwards, while the academic tracks last for three years and are intended to prepare pupils for university or college.

There are some important facts about the Norwegian school system that should be mentioned. First, compulsory education in Norway is free and all children over the age of six living in Norway for more than three months have both a right and a duty to attend.² Second, there is no ability tracking in primary or lower secondary education, and only 3.5 per cent of all pupils attend private schools.³ Third, school districts in Norway are fixed based on place of residence, and the general rule, which is strictly enforced, is that pupils attend the school in their district.⁴ Fourth, as shown in Bedard and Dhuey (2006) and Strøm (2004), repeating or skipping a grade is not common in the Norwegian school system, which means that most pupils start lower secondary education the year they turn thirteen.⁵

When starting compulsory education, children are assigned to classes that they typically attend for the entire duration of primary education, meaning that their peers remain the same throughout primary school. As they move up to lower secondary, most pupils change schools,⁶ and during this transformation, new classes are formed. Which class pupils are assigned to, both at primary and lower secondary school, should meet their need for social belonging and stability within the class, independent of pupil differences. Pupils or parents can express their wishes, but it is the school that has the final decision when it comes to allocation. According to the The Education Act (Opplæringslova) (1998) and the Ministry of Education and Research (2017), the assignment of pupils should strive to be as close to the aggregate population as possible and without consideration of ability, gender or ethnicity, meaning that the assignment should be random, but reflect the pupil population with regards to gender and other characteristics. This also applies for pupils receiving special education (Ministry of Education and Research, 2017). In my sample, I test whether assignment to a class with at least one low-achieving pupil is random within a given school.⁷

² Ref. the The Education Act (Opplæringslova) (1998) §2-1.

³ Ref. the The Education Act (Opplæringslova) (1998) §8-2 (1) and Statistics Norway (SSB).

⁴ Ref. the The Education Act (Opplæringslova) (1998) §8-1. If parents for some reason want their children to attend another school than the one assigned, they have to write a formal application to the local authority and there have to be available places in the school they want to be reassigned to, ref. the The Education Act (Opplæringslova) (1998) §8-1 (3). In addition, parents have to provide transportation to the new school.

⁵ Ref. the The Education Act (Opplæringslova) (1998) §2-1 (3) and (4).

⁶ Not all pupils change schools at this time. Some pupils in Norway attend integrated schools where they stay in the same school from first to tenth grade. About 25 per cent of Norwegian schools are integrated schools, located in smaller places, with fewer pupils, ref. Statistics Norway (SSB).

⁷ Ref. column 5 in Summary Statistics in Table 1.

¹ The Bad Apple model is explained in detail in studies by Hoxby and Weingarth (2005) and Lazear (2001).

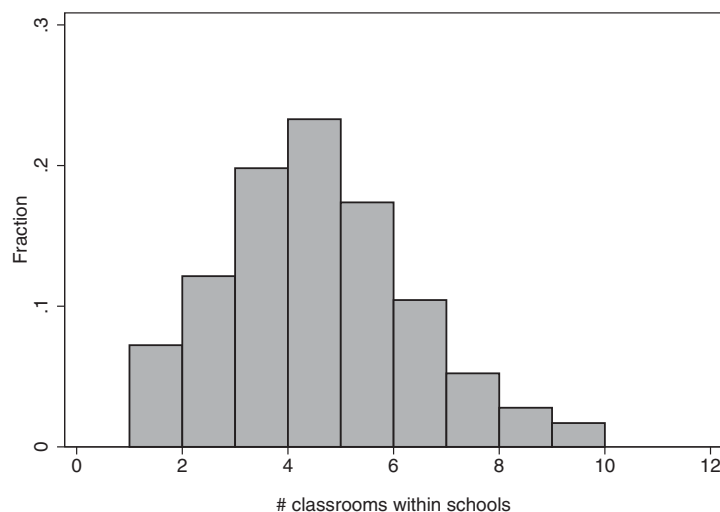


Fig. 1. Distribution of pupils. Notes: Fraction of pupils within different lower secondary (9th grade) schools based on number of classrooms.

The classes that pupils are assigned to, both in primary and lower secondary education, are the classes in which they receive all their education, meaning that their peers remain the same in all subjects, and that the only thing changing from subject to subject is the teacher.⁸ Figure 1 presents an overview of the number of classes within lower secondary schools in my sample, and how pupils are distributed.⁹

2.2. Inclusive education

The focus in the Norwegian school system is highly centred around equal opportunities to learn and inclusive education, and as early as the 1970s Norwegian politicians gave their support to a vision on adapted education (Boarini, 2009; Lillejord, 2015; Ogden, 2014). In 1992, the state special boarding schools were transformed or closed down as a part of the restructuring of the special needs education, and a reform within upper secondary education and training was launched in 1994 (Reform 94), where the main purpose was to safeguard a statutory right to attend upper secondary education for all pupils aged sixteen to nineteen (Ogden, 2014). The focus on inclusive education is still important in Norway, something that is strongly reflected in the *The Education Act (Opplæringslova)* (1998), with separate chapters on the right to receive special education and how to organise the pupils. The *The Education Act (Opplæringslova)* (1998, §8-2 (1)) states that pupils should be placed into classes or groups that meet their needs for social belonging, and that pupils should not, on a regular basis, be organised according to ability, gender or ethnicity. In the compulsory education, the focus is to provide education in accordance with the individual pupils' capabilities and needs in order to create equal opportunities for everybody, no matter where they live or what their socioeconomic status might be (Boarini, 2009). An important tool to be able to do this is special education.

Special education became an individual right for pupils in Norway in 1998, when the government implemented it as a part of the *The Education Act (Opplæringslova)* (1998, chapter 5). It states that pupils who not

sufficiently benefit from standard education may be entitled to special education or special language training. Special education could mean that the pupil works towards other learning goals than the rest of the class, that he or she is closely monitored and followed up by a teacher or assistant in a normal class setting, or that the pupil receives special equipment. What kind of special education a pupil receives depends on his or her needs. Pupils that do not have Norwegian as their first language may be entitled to special language training. Special language training continues until the pupil has gained sufficient language skills in Norwegian to be able to follow the standard education. The crucial factor for receiving special education or special language training is that the pupil is not able to benefit from the standard education.

In this respect, the definition of special educational needs (SEN) pupils in Norway is much closer to the definition of low-achievers, and somewhat different from that in many other countries. In most other countries, pupils actually need to have a diagnosis before receiving special education (OECD, 2012). Of the pupils defined as low-achievers in my sample, ref. section 3, the pupils I know to be SEN pupils¹⁰ make up about half of the observations, and just over half of the observations I have of SEN pupils are defined as low-achieving pupils.¹¹

Public spending on education in Norway is above the OECD average, with expenditures on educational institutions at levels at 6.2 per cent of GDP in 2014, mostly from public funds. In comparison, the OECD average is 5.2 per cent (OECD, 2017),¹² and the extra resources allocated to special needs children in the Norwegian school system are massive.

¹⁰ That is, pupils who are exempt from participating, and thus receiving special education in at least one of the three subjects - English, reading and maths - in 5th grade.

¹¹ A total of 1,876 of the 3,592 observations of low-achievers are of SEN pupils and 1,876 of the 3,305 observations of SEN pupils are defined as low-achievers. However, since I do not observe SEN pupils perfectly, but use a proxy for this, some of the 1,716 observations of non-SEN pupils defined as low-achievers might in fact be SEN pupils. See discussion under "Defining pupils who are exempt from participating in the national tests as low-achievers" in the robustness tests.

¹² Cumulative expenses per pupil by educational institution over the expected duration of primary and secondary studies amount to almost EUR 180,000, compared to the OECD average of just over EUR 110,000, and are exceeded only by Luxembourg (EUR 250,000) and Switzerland (EUR 180,000) (OECD, 2017).

⁸ According to the *The Education Act (Opplæringslova)* (1998) §8-2 (1), for parts of their education, pupils can be placed into smaller groups.

⁹ For instance, about 20 per cent of pupils attend schools with 3 classes and almost 25 per cent of pupils attend schools with 4 classes.

According to an OECD paper from 2009, pupils eligible for special language education receive 55–80 per cent more resources than the average pupil, and 65–130 per cent more resources are allocated to pupils with special needs than to other pupils in the Norwegian school system (Boarini, 2009).

According to the said OECD paper, Norwegian schools are quite successful when it comes to creating a safe and inclusive school environment, with pupils reporting low levels of bullying and that they generally enjoy going to school (Boarini, 2009). Between 2002 and 2004, only 0.4 per cent of Norwegian pupils in compulsory education were educated in special settings, compared to 5 per cent of pupils in the Czech Republic, 4.9 per cent of pupils in Flemish communities in Belgium and 4.8 per cent of German pupils (Eurydice, 2005; Ruijs and Peetsma, 2009). In addition, when it comes to equal opportunities, there are relatively few advantages to attending one school rather than another. Variation between schools in terms of average performance is among the lowest in the OECD countries, with only Finland and Iceland experiencing lower variation (Boarini, 2009).

Finally, people in Norway generally do not move in order to get their children into different schools, which is also the case for pupils in my sample. For example, of the pupils in my sample who started in a class at lower secondary school with at least one other pupil from their primary school, only one per cent were not in classes with anyone from primary school when attending 9th grade. This means that one per cent is the maximum amount of pupils who changed school between 8th and 9th grade of the pupils who started in the same school as their peers from primary school in my sample. Assuming pupils move together, this number could be larger, and the corresponding number when looking at pupils in classes with six or less pupils from primary school is only five per cent, meaning that only five per cent of the pupils could have moved. Of the pupils starting the same class in 8th grade, regardless of which primary school they attended, 97 per cent attend a class with at least one of the same peers in 9th grade, and about 90 per cent of the pupils in my sample attend classes with at least 10 peers from 8th grade.

3. Data description

I use registry data for the entire Norwegian population, collected and maintained by Statistics Norway. These include test scores from national tests, information about which school pupils attended, and individual characteristic information about pupils and their parents. In addition, as of 2011, the data contains class identifiers, making it possible to identify the exact class pupils are a part of, and thereby identifying their peers in a more accurate way than previously when the data only contained information about the school.

As described in section 4, the variables identifying class and school are crucial for my identification strategy, and because of this, I restrict my sample to include only pupils where these variables are recorded. In addition, I only keep pupils who participated or were exempt from participating in the national tests.¹³ Another crucial part of my identification strategy is that I need test scores both in 5th and 9th grade. I therefore exclude pupils for whom either one of these is missing. I also exclude pupils who do not have normal progression through school, since repeating or skipping a grade is not common in the Norwegian school system.¹⁴

When investigating peer effects, it is important that the group of interest has integrity as a peer group. This means that it is important

that we know that the group in question is in fact a group where members spend time together, and therefore it is reasonable to assume that they influence each other. One example of settings where this could be a problem is in neighbourhoods; even if people live close together, there is no way of knowing if they actually spend time together or even talk to each other, and therefore it is unknown whether they influence each other or not if they are not observed more closely. As mentioned in section 2, pupils in the Norwegian compulsory education attend the same group in all subjects in both primary and lower secondary school, and the only thing changing from subject to subject is the teacher, meaning that their peers remain the same throughout the school day, and because of this, I assume that in this setting there is no question with regards to integrity as a peer group. In order to get more accurate measures of peer effects and make sure the group in question in fact has integrity as a peer group, I exclude pupils who were in classes with less than 10 or more than 30 pupils. The reason for excluding these pupils is that the group variable in many of these cases can be wrongly recorded. In other words, they may be recorded for smaller groups within the class or whole cohorts within the school. This leaves a sample of 90,945 observations divided between reading and maths for 47,872 different pupils. When performing the regressions, I also drop the low-achievers themselves,¹⁵ leaving the final sample when estimating the effects to be 87,353 observations in the two subjects for 45,983 different pupils.

3.1. National tests¹⁶

The registry data used in this study contains test scores from national tests of pupils attending 5th, 8th and 9th grade. The purpose of the national tests is to provide knowledge about pupils' basic skills in English, reading and maths. The tests were performed for the first time in 2004/05, when pupils were tested in all three subjects in 5th and 8th grade. In 2010, the testing was broadened to also include pupils in 9th grade in the subjects reading and maths. The national tests are conducted during the autumn term, right after the pupils have started 5th, 8th and 9th grade. As a general rule, all pupils are to participate, but pupils requiring special education may be granted exemption if it is clear that the results from the test will not affect their future learning.¹⁷

The Norwegian Directorate for Education and Training is responsible for the development of the national tests, and do so in cooperation with professional communities within universities, colleges and national centres. The tests are electronic and externally graded, and consist of different texts, pictures and assignments with different degrees of difficulty, in order to provide information about pupils at all levels. The results are used by various parties. Teachers, for instance, use the results to follow up and adapt the education to their pupils, while local authorities and schools use the results to secure quality in the education. In addition, researchers can apply for and get the results to use them for scientific purposes.

3.2. Defining the treatment variable

Studies in the peer effect literature have suggested that including disadvantaged pupils in the mainstream classrooms has a negative effect on other pupils. Measures used include repeaters (Lavy et al., 2012a), pupils in the bottom quartile of the countrywide distribution (Hoxby and Weingarth, 2005), pupils from disadvantaged backgrounds (Stinebrickner and Stinebrickner, 2001), pupils exposed to domestic violence (Carrell et al., 2016; Carrell and Hoekstra, 2010), pupils with a psychiatric diagnosis or divorced or criminal parents (Kristoffersen et al., 2015), "Boys named Sue" (Figlio, 2007) and pupils in the bottom five per cent of the ability distribution (Lavy et al., 2012b).

¹³ Pupils can be exempted from participating in the national tests if they receive special education, ref. below describing national tests, section 2 and Appendix A describing who receives special education. However, I include these pupils, because as shown in Appendix Table A.1, these pupils most likely score below the ones that participated in the national tests in 9th grade, and are more likely to be exempted in 9th grade as well. Thus, these pupils are more likely to be the "true" low-achievers than pupils who participate in the national tests.

¹⁴ Ref. section 2.

¹⁵ The reason for excluding the low-achievers is that I want to investigate the effect on the other pupils, not the low-achieving pupils.

¹⁶ The Norwegian Directorate for Education and Training (2018)

¹⁷ Ref. Appendix A for more information about participation status.

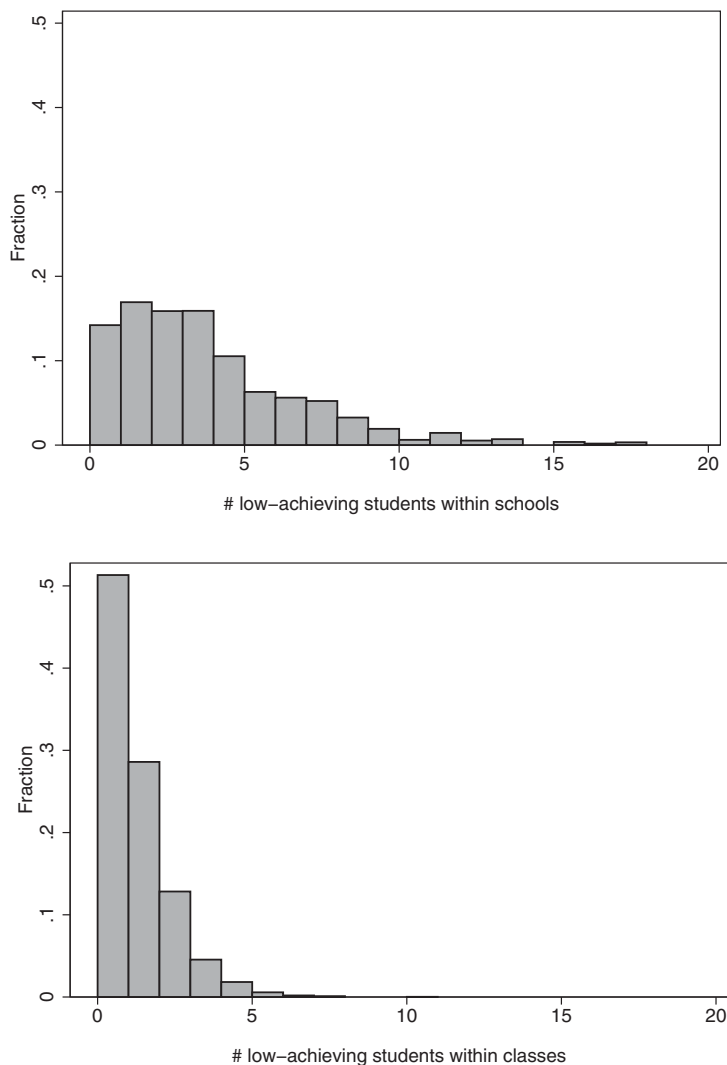


Fig. 2. Distribution of low-achievers. Notes: Fraction of pupils in schools with different number of low-achieving pupils in the first graph, and fraction of pupils exposed to different number of low-achieving pupils in the classroom in the second graph. Based on schools and classrooms in 9th grade.

Hanushek et al. (2002), on the other hand, find small, positive but insignificant estimates, when investigating the effect of being in a classroom with a disabled peer. Other literature investigating the effects of mainstream classrooms, the inclusive education literature, mostly focuses on SEN pupils when investigating effects on both the SEN pupils themselves and their peers (Kalambouka et al., 2007; Ruijs, 2017; Ruijs and Peetsma, 2009; Ruijs et al., 2010). Reviews from this literature report no negative effect (Kalambouka et al., 2007) and difficulties in drawing a clear conclusion (Ruijs and Peetsma, 2009) from existing literature.

In this study, I investigate whether there is evidence of the Bad Apple model described in studies by Lazear (2001) and Hoxby and Weingarth (2005) in the inclusive Norwegian school system. I do this

by investigating how pupils in the extreme lower tail of the achievement distribution, which I define as low-achieving pupils, affect their peers. Because I am able to make a national subject- and cohort-specific achievement distribution of the pupils based on test scores in the national tests in primary school 5th grade, which is before they start lower secondary, I can get around the reflection problem.¹⁸ I use this achievement distribution as a measure of pupil quality. I define low-achievers as pupils who score below the 10th percentile in the national subject- and cohort-specific achievement distribution in at least two of the three core subjects - English, reading and maths - in the national tests in 5th grade. These pupils are defined as low-achievers in all three subjects.

¹⁸ Explained in section 4.

Table 1
Summary statistics and balance test

	All pupils (1)	Control (2)	Treatment (3)	Difference (4)	Difference FE (5)
<i>Standardised test scores:</i>					
Reading 5th grade	0.000 (1.000)	0.032 (0.996)	-0.032 (1.003)	-0.064** (0.010)	-0.005 (0.012)
Maths 5th grade	0.000 (1.000)	0.044 (0.993)	-0.046 (1.005)	-0.091** (0.010)	-0.007 (0.012)
<i>Pupil Characteristics:</i>					
Male	0.501 (0.500)	0.503 (0.500)	0.500 (0.500)	-0.004 (0.003)	0.000 (0.004)
Immigrant	0.281 (0.449)	0.275 (0.446)	0.287 (0.452)	0.012** (0.003)	-0.001 (0.004)
<i>Parental Characteristics:</i>					
Mother's age	29.85 (5.05)	30.19 (5.01)	29.69 (5.08)	-0.32** (0.03)	0.00 (0.04)
Father's age	32.76 (5.88)	32.87 (5.85)	32.64 (5.92)	-0.24** (0.04)	-0.01 (0.05)
Mother's education	13.40 (3.56)	13.54 (3.54)	13.25 (3.58)	-0.28** (0.02)	-0.00 (0.03)
Father's education	13.05 (3.77)	13.19 (3.78)	12.90 (3.75)	-0.29** (0.03)	0.01 (0.03)
Parental income	1 063 (652)	1 103 (727)	1 023 (562)	-80** (4)	1 (5)
Observations	87 353	44 040	43 313	87 353	87 353

Notes: Mean statistics with standard deviations in parentheses in columns 1 to 3 and estimated coefficients with robust standard errors in the parentheses in columns 4 and 5. Summary statistics for pupils attending 9th grade in 2015. The first column shows summary statistics for all pupils in the sample, while the second and third columns show summary statistics for pupils not exposed to any low-achievers in 9th grade (control group) and pupils exposed to at least one low-achieving pupil in 9th grade (treatment group), respectively. The fourth column investigates if there are significant differences between the control and treatment groups without including any fixed effects, while the fifth column shows differences with fixed effect for school in 9th grade. Significance levels are indicated as follows: + significant at 10 per cent; * significant at 5 per cent; ** significant at 1 per cent. Parental age corresponds to the parent's age the year the child is born, while education is measured the year before the child is born. Both parental age and education are measured in years. Parental income is the income of both parents combined, measured in NOK and divided by 1000, and collected the year before the child is enrolled in 9th grade.

The reason for using this definition is to avoid defining someone as a low-achiever based on mistakes.¹⁹ In addition, I assume that if a pupil is in fact a low-achiever and belongs in the extreme lower tail of the achievement distribution, they are bound to score beneath the 10th percentile in at least two of the subjects, and thus by using this definition the data will provide a more accurate measure of low-achievers than using only one subject. The reason for not using all three subjects is that I found this to be too strict with regards to missing data on test scores. This gives 3,592 observations of low-achievers divided between 1,889 different pupils in the two subjects reading and maths (3.95 per cent of the sample). In Figure 2, I show how pupils are distributed in schools and classrooms with a different number of low-achieving peers in 9th grade.

The reason for using a 10 per cent threshold for defining low-achieving pupils is that this is equivalent to one low-achiever within each class potentially affecting the other pupils.²⁰ However, in the robustness tests, I include analyses using different thresholds for defining low-achieving peers.

An implication of defining low-achieving pupils as described above is that within a cohort, low-achieving pupils will always exist, and pupils at a certain achievement level might be defined as a low-achiever one year, but not the following. For example, in a particularly good cohort a pupil could be labelled as low-achieving, but might not be if he or she were to start school with the pupils starting the year after.

In my main regression analysis, I investigate the effect of these pupils on their peers by creating a dummy variable for being exposed to at least one low-achieving peer in the class in 9th grade, and use this as the independent variable of interest. If I find that the presence of at least one low-achieving pupil in the classroom has negative effects on their peers, this indicates that there is evidence of the Bad Apple model (Hoxby and Weingarth, 2005; Lazear, 2001).

This is a slightly different approach from most studies in the peer effect literature, which use the share of low-achievers in the classroom as the treatment variable. The reason for using a different variable is that I am interested in what happens with the achievement if pupils are exposed to that one potentially disruptive peer who interrupts the education compared to pupils who are not exposed to any low-achieving peers, ref. the Bad Apple model (Hoxby and Weingarth, 2005; Lazear, 2001; Sacerdote, 2011). It is more in line with the approach used in the inclusive education literature, where the researchers look at differences in inclusive versus non-inclusive settings. However, I also include a specification analysis where I investigate the effects using the continuous variable share of low-achievers in the classroom, and look at being exposed to more low-achieving peers, in order to be able to investigate if this changes the results.

3.3. Summary statistics

Summary statistics for my main independent variables are presented in Table 1. The statistics show that about 50 per cent of the pupils in my sample are male and about 30 per cent come from immigrant backgrounds. Mothers are about 30 years old when giving birth, while fathers are a little older when the child is born (almost 33 years). The mothers of the pupils in my sample have about 13.5 years of education, while

¹⁹ For example, SEN pupils who are not low-achievers, but have difficulties with the testing situation in one of the subjects (e.g. pupils with autism, ADHD, etc.).

²⁰ Classes in my sample on average consist of 22 pupils, so 3.95 per cent of the observations corresponds to one pupil in each class.

Table 2
Test of identifying assumption

	Dependent variable						
	Male (1)	Immigrant (2)	Mother's age when the child is born (3)	Father's age when the child is born (4)	Mother's education (5)	Father's education (6)	Parental income (7)
<i>Panel A. Both subjects</i>							
Having "low-achieving" peer(s)	0.001 (0.005)	-0.002 (0.004)	0.024 (0.047)	0.038 (0.055)	0.028 (0.034)	0.025 (0.036)	7 207 (5 533)
<i>Panel B. Reading</i>							
Having "low-achieving" peer(s)	0.004 (0.007)	-0.001 (0.006)	0.037 (0.066)	0.059 (0.077)	0.035 (0.049)	0.033 (0.050)	8 211 (7 799)
<i>Panel C. Maths</i>							
Having "low-achieving" peer(s)	-0.001 (0.007)	-0.003 (0.006)	0.011 (0.066)	0.018 (0.078)	0.020 (0.049)	0.018 (0.050)	6 208 (7 852)

Notes: Fixed effects regressions, with standard errors clustered on class in 5th grade and school in 9th grade in parentheses. Significance levels are indicated as follows: + significant at 10 per cent; * significant at 5 per cent; ** significant at 1 per cent. Dependent variable is male, immigrant, mother's age when the child is born, father's age when the child is born, mother's education, father's education and parental income in columns 1 to 7, respectively, and the coefficient of interest is the coefficient on the dummy for having low-achieving peer(s) in 9th grade. Coefficients are from separate regressions, where I include fixed effects for class in 5th grade and school in 9th grade. Panel A shows the effect in both subjects combined, while Panels B and C show the results separately for reading and maths, respectively.

the fathers have a little less (about 13 years). The parental income is just above 1 MNOK.²¹

When dividing the sample into treatment (those exposed to at least one low-achieving peer in 9th grade) and control (those not exposed to any low-achieving peers in 9th grade), and investigating the differences between the two, in column four I find that pupils in the treatment group on average seem to have somewhat lower test scores in both reading and maths in 5th grade. In addition, I find that there are more girls and more immigrants in the treatment group. Parents of pupils in the treatment group tend to be a little younger, have a bit less education and earn slightly less. The differences are small, but most of them are significant, and in the regression I include them as control variables in order to control for the differences presented here. When adding fixed effects for school in 9th grade, in column five, these differences disappear, meaning that within a given school pupils are randomly assigned to classes with at least one low-achieving peer. When I include fixed effects for pupils attending the same class in 5th grade and the same school in 9th grade, as shown in Table 2, the treatment remains balanced.

4. Empirical strategy

Models in the peer effects literature struggle with three main problems that could potentially arise when trying to identify the effects, thereby making the estimated coefficients biased (Angrist, 2014; Feld and Zölitz, 2017; Manski, 1993). The first one is the reflection (or multiplier) problem, which an entire study has been devoted to (Manski, 1993), reflecting the magnitude of the problem. In his paper, Manski (1993, p. 531) defines the reflection problem as follows: "The reflection problem that arises when a researcher observing the distribution of behaviour in a population tries to infer whether the average behaviour in some group influences the behaviour of the individuals that comprise the group," which could be summarised in his question: "Does the mirror image cause the person's movements or reflect them?" (Manski, 1993, p. 532).

The second problem, known as the selection problem, occurs if individuals, when choosing their peers, sort themselves into groups based on specific characteristics, like ability (Lavy et al., 2012a). The group's outcome could then appear to cause the pupil's outcome, when in fact it is the other way around. The third problem is the measurement error or omitted variable problem, which occurs because a deciding factor of the

pupil's achievement is either badly measured or omitted, and thereby included as part of the error term (Hoxby and Weingarth, 2005).

In order to get around the reflection problem, an essential requirement is that the determination of pupil quality is based on outcomes prior to the change of school when starting lower secondary and being assigned to a new class. The structure of the data, presented in section 3, provides me with enough information to be able to do this, by linking pupils' 5th grade test scores (taken in primary school) with their 9th grade test scores (taken in lower secondary school). In addition, by tracking pupils over two time periods, I am able to identify which peers are new, i.e. come from a different primary school class, and which peers are old, i.e. come from the same primary school class. On average in my sample, about 75 per cent of pupils' peers in secondary school did not come from the same primary school class, and because of this, I assume that their test scores from 5th grade could not have been affected by these peers. I therefore use the 5th grade test scores to identify two groups of pupils; pupils who are low-achievers and other pupils. Low-achievers are, as mentioned in section 3, defined as pupils below the 10th percentile of the national subject- and cohort-specific achievement distribution in 5th grade in at least two of the three subjects. The rest of the pupils are defined as other pupils.

When investigating the effect of having at least one low-achieving peer in 9th grade on a pupil's test scores, I use a standard education production function model. This means that I estimate outcomes, e.g. test scores, as a function of school, class and individual characteristics, in the following fixed effects regression:

$$y_{icst} = \phi_1 \alpha_i + \phi_2 y_{i(c-1)s(t-1)} + \delta_{(c-1)(t-1)} \times \gamma_{st} + \beta_1 D_{icst}^l + \varepsilon_{icst} \quad (1)$$

where y_{icst} is the achievement measure of pupil i , in class c , in school s at time t . In this paper, I use test scores from national tests in 9th grade as the achievement measure. α_i represents controls for individual characteristics (including parental characteristics), $y_{i(c-1)s(t-1)}$ represents control for 5th grade test scores, the expression $\delta_{c-1,t-1} \times \gamma_{st}$ represents class in 5th grade and school in 9th grade fixed effects, i.e. fixed effects for pupils attending the same class in 5th grade and the same school in 9th grade. D_{icst}^l is a dummy variable taking the value one if the pupil is exposed to at least one low-achieving peer in the class in 9th grade, and zero if not, and ε_{icst} is an error term.

In particular, I use the redistribution of pupils when starting lower secondary school, and thereby the possible change in being exposed to low-achieving peers to investigate if the exposure to at least one low-achieving pupil in the classroom has any effect on a pupil's achievement in the national tests in 9th grade. My identifying assumptions are that

²¹ Which corresponds to about EUR 115 000.

there is variation in the exposure to low-achievers within schools and that any individual characteristics of pupils starting the same school and who have previously attended the same class are independent of the likelihood of ending up in a class with at least one low-achieving pupil in lower secondary school.

This strategy deals with the potential selection problem by assuming pupils' peer groups within a given school are randomly assigned,²² and including fixed effects and control variables, allowing me to control for individual and parental characteristics. The potential problem concerning measurement error, or omitted variables, is impossible to exclude completely, but by using the exact same test for all pupils within the same subject and cohort (with multiple-choice questions, guidance for examinations on open questions and external grading), I am able to exclude measurement errors with regards to examinations and instructions. Based on the assumption that the problems presented in the peer effects literature are taken care of, equation (1) provides me with an unbiased estimate of the coefficient of interest, β_1 , which captures the effect of the presence of at least one low-achieving peer.

In addition to the effect on the full sample, I want to examine whether there are differences across different subjects, gender, socioeconomic status and if the pupils' own position in the achievement distribution affect how they respond to being exposed to at least one low-achieving peer. In order to investigate this, I use subsample analyses based on subject, gender, parental education and earnings, and pupils' own position in the achievement distribution, both with regards to all low-achievers in the subsample analysis, and with different independent variables based on the gender of the low-achiever and the pupil in the specification analysis. In order to investigate if the effects change when using the continuous variable share of low-achievers in the classroom and being exposed to more low-achieving peers, I also include specification analyses where I use this as the independent variables of interest.

5. Empirical results

This section first presents a test of my identifying assumptions, followed by my main findings and a range of heterogeneity, specification and robustness analyses.

5.1. Test of identifying assumptions

As mentioned, my identifying assumptions are that there is variation in the exposure to low-achievers within schools and that any individual characteristics of pupils starting the same school and who have previously attended the same class are independent of the likelihood of ending up in a class with at least one low-achieving pupil in lower secondary school. In Figure 3, I show the fraction of pupils attending the same class in 5th grade and the same school in 9th grade who are exposed to at least one low-achieving peer. The graph shows that approximately 60 per cent of the pupils in my sample experience variation in the exposure to low-achieving peers within their group,²³ and the fraction of exposed pupils varies from 4 to 96 per cent. Among 50 per cent of the pupils in my sample, the fraction of exposed pupils varies between 20 and 80 per cent.²⁴ In order to get an understanding of the number of pupils in these groups, the second graph in Figure 3 shows the distribution of pupils across the groups attending the same class in 5th

grade and the same school in 9th grade. The final graph in Figure 3 illustrates the same as the first, but on group level.²⁵

With regards to the likelihood of ending up in a class with at least one low-achieving pupil, I test whether pupils are randomly assigned to classes with regards to individual and parental characteristics in Table 2. This is done by performing regressions with individual and parental characteristics as the dependent variable and the dummy for being exposed to at least one low-achiever in 9th grade as the independent variable of interest, with fixed effect for attending the same school in lower secondary and the same class in primary school. All of the estimated coefficients in Table 2 are very small, and none of them are significantly different from zero, which means that in my sample there is random assignment to classes with at least one low-achieving pupil with regards to the characteristics investigated here. Based on the graphs and analysis presented here I assume that my identifying assumptions hold.

5.2. Main results

Table 3 shows the results from several specifications investigating the effect on the full sample of pupils. In all the specifications, I first present the estimated effects in both subjects combined, in Panel A, before dividing the sample based on the subjects reading and maths, in Panel B and C, respectively. In the first column I present the estimated effects from the OLS regression with controls for individual and parental characteristics. In columns two and three I first add fixed effects for class in 5th grade and then school in 9th grade. In my preferred specification, presented in column four, I group pupils attending the same class in 5th grade and the same school in 9th grade and use this as the fixed effects. The results from the preferred model show that being exposed to at least one low-achiever in the classroom in 9th grade reduces performance in the national tests by 1.2 per cent of a standard deviation. However, this is only significant at a 10 per cent level, and no longer significant when dividing the sample based on subject.

5.3. Heterogeneity

The effects presented in the main analysis suggest that there is no effect on pupils not defined as low-achievers when being exposed to low-achieving peers. However, because of fundamental differences in interest and behaviour, peer effects might play out very differently for different pupils, and even though pupils on average are not affected, some groups of pupils may experience effects from being exposed to low-achieving peers. Much of the evidence in the peer effect literature indicates that peer effects are heterogeneous and non-linear in form, as opposed to homogeneous and linear as previously assumed (Carrell et al., 2009; Fruehwirth, 2014; Gibbons and Telhaj, 2008).²⁶ Because of this, I now expand the analysis to also include several subsample analyses in order to investigate whether the effects are different for different groups of pupils.

In the first subsample analysis, I investigate the effects in both subjects combined and separately based on gender,²⁷ presented in Table 4. These analyses show that there does not seem to be any effect on either boys or girls from the presence of at least one low-achieving peer in the classroom, as all the estimated coefficients are very small and none of them are significantly different from zero.

Previous research has also looked at differences with respect to socioeconomic background, which I do in my second subsample analysis,

²² Based on the fact that pupils are randomly assigned to classes within schools and not according to ability or other unobserved characteristics in primary and lower secondary school in Norway, ref. section 3.

²³ About 20 per cent of the pupils are in groups where none of the pupils are exposed to low-achieving peers (the first bar in Figure 3) and about 20 per cent of the pupils are in groups where all the pupils are exposed to low-achieving peers (the last bar in Figure 3).

²⁴ In Appendix B, I show the distribution of pupils in classes with different number of low-achieving peers within schools with more than one low-achiever.

²⁵ Because there is no variation in the exposure to low-achieving pupils in the groups that only have one pupil (the first bar in the second graph), these groups are excluded in the third graph.

²⁶ Ref. also previous studies by Feld and Zölitz (2017); Hanushek et al. (2003); Hoxby and Weingarth (2005); Lavy et al. (2012b), etc. investigating heterogeneity with regards to peer effects.

²⁷ Low-achievers can be of both genders in this analysis.

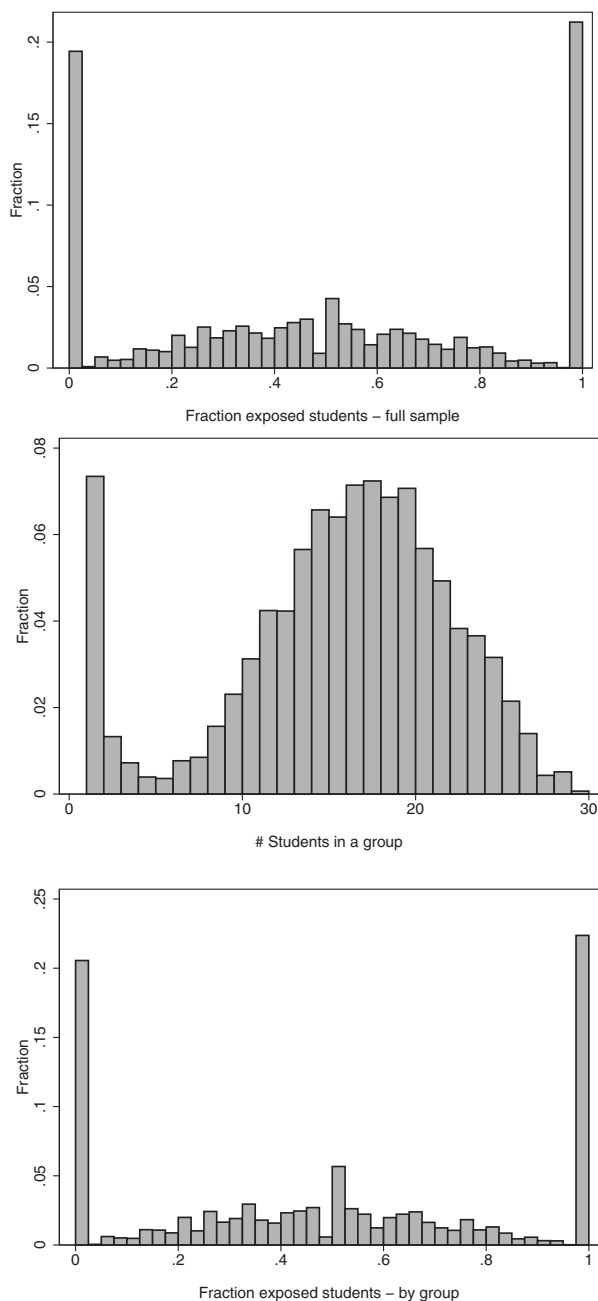


Fig. 3. Fraction of pupils attending the same class in 5th grade and the same school in 9th grade exposed to at least one low-achieving peer. Notes: The first graph shows the fraction of pupils exposed to at least one low-achieving pupil in the group of pupils attending the same class in 5th grade and the same school in 9th grade on the individual level. The second graph illustrates the number of pupils within each group, and how pupils are distributed based on this. The third graph shows the fraction of exposed pupils on group level (because there is no variation among pupils not attending the same school as any other pupils from their 5th grade class in 9th grade, pupils from the first bar in the second graph are excluded when looking at the group level in the third graph).

Table 3
Effect of low-achieving peer(s)

	(1)	(2)	(3)	(4)
<i>Panel A. Both subjects</i>				
Having “low-achieving” peer(s) in 9th grade	-0.009 ⁺ (0.005)	-0.014 [*] (0.006)	-0.012 ⁺ (0.006)	-0.012 ⁺ (0.007)
Having “low-achieving” peer(s) in 5th grade	0.044 ^{**} (0.005)			
<i>Panel B. Reading</i>				
Having “low-achieving” peer(s) in 9th grade	-0.010 (0.007)	-0.010 (0.009)	-0.011 (0.009)	-0.009 (0.010)
Having “low-achieving” peer(s) in 5th grade	0.028 ^{**} (0.008)			
<i>Panel C. Maths</i>				
Having “low-achieving” peer(s) in 9th grade	-0.007 (0.007)	-0.016 [*] (0.008)	-0.014 (0.009)	-0.014 (0.009)
Having “low-achieving” peer(s) in 5th grade	0.060 ^{**} (0.007)			
<i>Control variables:</i>				
Individual characteristics	X	X	X	X
Parental characteristics	X	X	X	X
<i>Fixed effects:</i>				
Class in 5th grade		X	X	
School in 9th grade			X	
Interaction class in 5th grade and school in 9th grade				X

Notes: OLS and fixed effects regressions, with robust standard errors in parentheses in the first three models and standard errors clustered on class in 5th grade and school in 9th grade in parentheses in model four (4). Significance levels are indicated as follows: + significant at 10 per cent; * significant at 5 per cent; ** significant at 1 per cent. Dependent variable is standardised test scores in national tests in 9th grade, and the coefficient of interest is the coefficient on the dummy for having low-achieving peer(s) in 9th grade. The coefficient on the dummy for having low-achieving peer(s) in 5th grade is also included in order to show that this disappears when fixed effects on class in 5th grade are included. The first model (1) is the OLS regression looking only at the relationship between having low-achieving peer(s) in 9th grade on a pupil's (standardised) test scores, controlling for having low-achieving peer(s) in 5th grade, cohort and individual and parental characteristics. In the second model (2), fixed effects for class in 5th grade are added, and in the third model (3) I add fixed effects for school in 9th grade. In model four (4), I group pupils attending the same class in 5th grade and the same school in 9th grade, thereby using an interaction term. Individual characteristics include pupils' standardised test scores in 5th grade and dummies for being male and an immigrant. Parental characteristics include parent's age when the child is born, parent's years of education and parental income. Panel A shows the estimated effect in both subjects combined, while Panels B and C show the effects in reading and maths, respectively.

Table 4
Subsample analysis by subject and gender

	Gender		Observations
	Boys	Girls	
<i>Panel A. Both subjects</i>			
Having “low-achieving” peer(s)	-0.008 (0.010)	-0.010 (0.009)	87 353
<i>Panel B. Reading</i>			
Having “low-achieving” peer(s)	-0.004 (0.016)	-0.006 (0.013)	43 707
<i>Panel C. Maths</i>			
Having “low-achieving” peer(s)	-0.011 (0.014)	-0.014 (0.013)	43 646
Observations	43 807	43 546	

Notes: Fixed effects regressions for boys and girls in both subjects combined and separately, with standard errors clustered on class in 5th grade and school in 9th grade in parentheses. Significance levels are indicated as follows: + significant at 10 per cent; * significant at 5 per cent; ** significant at 1 per cent. Dependent variable is standardised test scores in national tests in 9th grade, and the coefficient of interest is the coefficient on the dummy for having low-achieving peer(s) in 9th grade. Coefficients are from separate regressions, based on model four (4) in Table 3, where I include fixed effects for the interaction between class in 5th grade and school in 9th grade, and controls for individual and parental characteristics. The first column represents boys and the second column represents girls. The low-achievers are not divided by gender. In Panel A, I look at both subjects combined, and in Panels B and C I look only at test scores in reading and maths, respectively. The “Observations” column and row include the number of observations within the different categories.

with parental education as a proxy for socioeconomic background, presented in Table 5.²⁸ When doing this I find significant negative estimates for pupils whose mother or father have a university degree when looking at the effects in both subjects combined, for pupils whose mothers have a university degree when investigating the effects separately in reading, and for pupils whose fathers have a university degree when investigating the effects separately in maths. The effects, however, are very small with negative effects between 2 and 3 per cent of a standard deviation, and only significant at a 10 per cent level.

In the last subsample analysis, I also investigate the effects based on pupils' own performance, based on their position in the national subject- and cohort-specific achievement distribution in 5th grade, presented in Table 6. In this analysis I find somewhat larger estimates than in the previous analyses, with significant estimates ranging from negative effects of 3.5 to 6.3 per cent of a standard deviation for different groups of pupils. However, in real numbers, the largest of these estimates corresponds to 0.5 lower test scores for the pupils in this group with a mean test score of 27.5 on a scale ranging from 0 to 44.

5.4. Specification analysis

5.4.1. Share of low-achieving pupils

As mentioned, other studies in the peer effect literature have used a slightly different approach when investigating the effects, by using the share of disadvantaged or low-achieving pupils. In this specification analysis, I change the independent variable to the share of low-achieving

²⁸ And by using parental income as a proxy in Appendix C.

Table 5
Subsample analysis by subject and parental education

	Mother's education			Father's education		
	Not completed upper secondary school	Completed upper secondary school	University Degree	Not completed upper secondary school	Completed upper secondary school	University Degree
<i>Panel A. Both subjects</i>						
Having "low-achieving" peer(s)	-0.021 (0.016)	-0.009 (0.013)	-0.020* (0.010)	-0.010 (0.016)	-0.013 (0.011)	-0.021* (0.012)
<i>Panel B. Reading</i>						
Having "low-achieving" peer(s)	-0.010 (0.024)	-0.012 (0.018)	-0.027* (0.015)	-0.008 (0.024)	-0.013 (0.016)	-0.013 (0.017)
<i>Panel C. Maths</i>						
Having "low-achieving" peer(s)	-0.032 (0.022)	-0.006 (0.017)	-0.011 (0.014)	-0.009 (0.020)	-0.014 (0.014)	-0.029* (0.015)

Notes: Fixed effects regressions for subsamples of pupils in both subjects combined and separately based on parental education, with standard errors clustered on class in 5th grade and school in 9th grade in parentheses. Significance levels are indicated as follows: + significant at 10 per cent; * significant at 5 per cent; ** significant at 1 per cent. Dependent variable is standardised test scores in national tests in 9th grade, and the coefficient of interest is the coefficient on the dummy for having low-achieving peer(s) in 9th grade. Coefficients are from separate regressions, based on model four (4) in Table 3, where I include fixed effects for the interaction between class in 5th grade and school in 9th grade, and controls for individual and parental characteristics. Parental education is specified in columns 1 to 6, with mother's education in columns 1 to 3 and father's education in columns 4 to 6. In Panel A, I look at both subjects combined, and in Panels B and C I look only at test scores in reading and maths, respectively.

Table 6
Subsample analysis by subject and pupils' own position in the achievement distribution

	Pupils' own position in the achievement distribution				
	Below 20th percentile	20th to 40th percentile	40th to 60th percentile	60th to 80th percentile	Above 80th percentile
<i>Panel A. Both subjects</i>					
Having "low-achieving" peer(s)	-0.015 (0.027)	-0.035* (0.019)	0.011 (0.018)	-0.017 (0.015)	-0.003 (0.013)
<i>Panel B. Reading</i>					
Having "low-achieving" peer(s)	0.031 (0.042)	-0.063* (0.027)	0.039 (0.027)	-0.003 (0.022)	-0.005 (0.018)
<i>Panel C. Maths</i>					
Having "low-achieving" peer(s)	-0.058* (0.035)	-0.005 (0.026)	-0.014 (0.024)	-0.032 (0.021)	-0.002 (0.019)

Notes: Fixed effects regressions for subsamples of pupils in both subjects combined and separately based on pupils' own position in the achievement distribution, with standard errors clustered on class in 5th grade and school in 9th grade in parentheses. Significance levels are indicated as follows: + significant at 10 per cent; * significant at 5 per cent; ** significant at 1 per cent. Dependent variable is standardised test scores in national tests in 9th grade, and the coefficient of interest is the coefficient on the dummy for having low-achieving peer(s) in 9th grade. Coefficients are from separate regressions, based on model four (4) in Table 3, where I include fixed effects for the interaction between class in 5th grade and school in 9th grade, and controls for individual and parental characteristics. Pupils' own position in the achievement distribution is specified in columns 1 to 5, with the lowest achieving pupils in column 1 and the highest achieving pupils in column 5. In Panel A, I look at both subjects combined, and in Panels B and C I look only at test scores in reading and maths, respectively.

peers instead of using a dummy for being exposed to at least one low-achiever, making D_{cst}^l presented in equation (1) in section 4 a continuous variable taking on values between 0 and 1 instead of a dummy variable, and find similar results as presented above. The estimated coefficients are presented in Tables 7, 8 and 9. Using my most comprehensive approach (fixed effects for the interaction between class in 5th grade and school in 9th grade) there is only one significant estimate, a positive effect of 39.1 per cent of a standard deviation for pupils above the 80th percentile in the national subject- and cohort-specific achievement distribution in maths. This measure, however, does not make much sense since a 100 per cent increase in the share of low-achieving peers never happens (meaning that all peers of a pupil have to be changed from only pupils defined as not low-achieving, to only low-achieving pupils). On average, classes in my sample consist of 22 pupils, which means that the average pupil has 21 peers. Changing one of these peers to a low-achieving pupil²⁹ corresponds to about a 5 per cent increase in the proportion of low-achieving peers. Based on the largest estimated effect, this would correspond to a 1.9 per cent standard deviation increase in

test scores ($0.391 * 0.05$), which is in line with the estimated coefficients found in the corresponding analysis presented above.

5.4.2. More low-achieving peers

One can argue that in order to affect other pupils, the low-achievers have to be more than one. In order to investigate this, I have run several regressions with exposure to one, two, three, four, five, six and seven low-achievers, respectively, and the estimated effects are presented in Figure 4, with the effect on all pupils in the first graph, boys in the second and girls in the third. As shown in the graph, there does not seem to be any effect before having five low-achieving pupils in the classroom, and then the effect seems to be negative, but for pupils exposed to six low-achieving peers the effect is positive. The total number of pupils exposed to five, six or seven low-achieving peers in the classroom in lower secondary school is just under 700. Thus, even though the effect is significantly different from zero for some of these pupils, there are very few pupils in these samples.

5.4.3. Gender differences

As briefly mentioned, peer effects might play out very differently for different pupils, and because of this I also want to investigate whether

²⁹ From a pupil defined as other pupil.

Table 7
Specification analysis using continuous variable - subject and gender

	All pupils	Gender	
		Boys	Girls
<i>Panel A. Both subjects</i>			
Fraction “low-achieving” peer(s)	-0.100 (0.075)	-0.001 (0.116)	-0.142 (0.103)
<i>Panel C. Reading</i>			
Fraction “low-achieving” peer(s)	-0.113 (0.110)	-0.030 (0.174)	-0.165 (0.144)
<i>Panel D. Maths</i>			
Fraction “low-achieving” peer(s)	-0.077 (0.102)	0.036 (0.152)	-0.115 (0.146)

Notes: Fixed effects regressions for all pupils in both subjects combined and separately based on subject and gender, with standard errors clustered on class in 5th grade and school in 9th grade in parentheses. Significance levels are indicated as follows: + significant at 10 per cent; * significant at 5 per cent; ** significant at 1 per cent. Dependent variable is standardised test scores in national tests in 9th grade, and the coefficient of interest is the coefficient on the share of low-achieving peers in the classroom in 9th grade. Coefficients are from separate regressions, based on equation (1) in section 4 where the dummy has been changed to a continuous variable taking values between 0 and 1, where I include fixed effects for the interaction between class in 5th grade and school in 9th grade, and controls for individual and parental characteristics. The first column represents all pupils, the second represents boys and the third column represents girls. In Panel A, I look at both subjects combined, and in Panels B and C I look only at test scores in reading and maths, respectively.

boys and girls are affected differently depending on their socioeconomic status or own achievement. The estimated coefficients are presented in Appendix Tables D.1, D.2, D.3 and D.4. As in the previous analyses, I do find some significant estimates scattered around, but with the largest one being a negative effect of 10.1 per cent of a standard deviation for boys below the 20th percentile in the achievement distribution in maths, the effects are very small in size and do not correspond to any “real” changes in test scores of the affected pupils.³⁰

³⁰ In real numbers, this corresponds to a 1 point lower test score for exposed pupils in this group with a mean test score of 20 on a scale ranging from 0 to 50.

In addition, because I am measuring the effects in lower secondary school, a time in children’s lives when there is a lot happening on a personal level which could affect who influences them at school, I also include a specification analysis where I divide the low-achieving pupils into boys and girls. That is, I include separate regressions where I divide the treatment variable based on gender, making the following equation the regression of interest:

$$y_{icst} = \phi_1 \alpha_i + \phi_2 y_{i(c-1)s(t-1)} + \delta_{(c-1)(t-1)} \times y_{st} + \beta_1 D_{cst}^{lb} + \beta_2 D_{cst}^{lg} + \varepsilon_{icst} \quad (2)$$

where the coefficients of interest are β_1 and β_2 , which captures the effect of the presence of at least one low-achieving boy (D_{cst}^{lb}) and one low-achieving girl (D_{cst}^{lg}), respectively. The estimated coefficients are presented in Appendix Tables D.5, D.6 and D.7. In these analyses, the estimates remain about the same as in previous analyses, shifting around zero, with some of them being significantly different from zero. However, all the estimated coefficients are very small in size, and there is no clear pattern as to any group of pupils being more or less influenced by one kind of low-achiever.

In the last set of specifications concerning gender, I combine the previous two specifications and look at the gender of both the pupil and the low-achiever. In order to do this, I use the same approach as presented above and investigate the effects in different groups of boys and girls depending on their socioeconomic status or own achievement, presented in Appendix Tables D.8, D.9, D.10 and D.11. By doing this, I get the same results as the other analysis; it does not seem as though pupils are affected by the presence of at least one low-achieving pupil in the classroom, whether boy or girl. The largest effect on boys seems to stem from low-achieving girls in reading on boys between the 40th and the 60th percentile, with a positive estimate of 13.6 per cent of a standard deviation. Meanwhile, the greatest effect on girls seems to stem from low-achieving girls in reading for the girls below the 20th percentile in the national subject- and cohort-specific achievement distribution, with a positive effect of 17.3 per cent of a standard deviation. Even if the estimated effects presented here are larger than in the previous analyses, in real numbers this corresponds to 1 point higher test scores in reading for exposed boys between the 40th and 60th percentile in the achievement distribution with a mean test score of 30 points on a scale ranging from 0 to 44. For girls, the positive effect of 17.3 per cent of a standard deviation corresponds to about 1.5 points higher test scores for exposed pupils below the 20th percentile in the achievement distribution with a mean test score of 22.5 on a scale ranging from 0 to 44. In addition, the

Table 8
Specification analysis using continuous variable - subject and parental education

	Mother’s education			Father’s education		
	Not completed upper secondary school	Completed upper secondary school	University Degree	Not completed upper secondary school	Completed upper secondary school	University Degree
<i>Panel A. Both subjects</i>						
Fraction “low-achieving” peer(s)	-0.220 (0.161)	-0.040 (0.130)	-0.054 (0.117)	0.024 (0.151)	-0.186 (0.116)	-0.180 (0.143)
<i>Panel C. Reading</i>						
Fraction “low-achieving” peer(s)	-0.212 (0.241)	-0.110 (0.192)	-0.155 (0.164)	0.027 (0.229)	-0.216 (0.173)	-0.164 (0.211)
<i>Panel D. Maths</i>						
Fraction “low-achieving” peer(s)	-0.212 (0.208)	0.030 (0.175)	0.058 (0.165)	0.037 (0.194)	-0.147 (0.154)	-0.203 (0.192)

Notes: Fixed effects regressions for subsamples of pupils in both subjects combined and separately based on subject and parental education, with standard errors clustered on class in 5th grade and school in 9th grade in parentheses. Significance levels are indicated as follows: + significant at 10 per cent; * significant at 5 per cent; ** significant at 1 per cent. Dependent variable is standardised test scores in national tests in 9th grade, and the coefficient of interest is the coefficient on the share of low-achieving peers in the classroom in 9th grade. Coefficients are from separate regressions, based on equation (1) in section 4 where the dummy has been changed to a continuous variable taking values between 0 and 1, where I include fixed effects for the interaction between class in 5th grade and school in 9th grade, and controls for individual and parental characteristics. Parental education is specified in columns 1 to 6, with mother’s education in columns 1 to 3 and father’s education in columns 4 to 6. In Panel A, I look at both subjects combined, and in Panels B and C I look only at test scores in reading and maths, respectively.

Table 9
Specification analysis using continuous variable - subject and pupils' own position in the achievement distribution

	Pupils' own position in the achievement distribution				
	Below 20th percentile	20th to 40th percentile	40th to 60th percentile	60th to 80th percentile	Above 80th percentile
<i>Panel A. Both subjects</i>					
Fraction "low-achieving" peer(s)	-0.054 (0.255)	-0.215 (0.197)	-0.046 (0.197)	-0.059 (0.171)	0.179 (0.146)
<i>Panel C. Reading</i>					
Fraction "low-achieving" peer(s)	0.216 (0.391)	-0.285 (0.279)	0.233 (0.277)	0.099 (0.239)	-0.077 (0.207)
<i>Panel D. Maths</i>					
Fraction "low-achieving" peer(s)	-0.278 (0.326)	-0.108 (0.274)	-0.299 (0.275)	-0.250 (0.240)	0.391+ (0.202)

Notes: Fixed effects regressions for subsamples of pupils in both subjects combined and separately based on subject and pupils' own position in the achievement distribution, with standard errors clustered on class in 5th grade and school in 9th grade in parentheses. Significance levels are indicated as follows: + significant at 10 per cent; * significant at 5 per cent; ** significant at 1 per cent. Dependent variable is standardised test scores in national tests in 9th grade, and the coefficient of interest is the coefficient on the share of low-achieving peers in the classroom in 9th grade. Coefficients are from separate regressions, based on equation (1) in section 4 where the dummy has been changed to a continuous variable taking values between 0 and 1, where I include fixed effects for the interaction between class in 5th grade and school in 9th grade, and controls for individual and parental characteristics. Pupils' own position in the achievement distribution is specified in columns 1 to 5, with the lowest achieving pupils in column 1 and the highest achieving pupils in column 5. In Panel A, I look at both subjects combined, and in Panels B and C I look only at test scores in reading and maths, respectively.

significant estimates do not reveal any clear patterns as to who is more or less affected by whom.

5.5. Robustness tests

5.5.1. Using different thresholds for defining low-achieving pupils

So far, I have used a threshold of 10 per cent in at least two of the three subjects when defining low-achieving pupils, thus defining about 4 per cent of the observations to be of low-achievers. I could have chosen different thresholds, potentially affecting my results. In the tables in Appendix E, I look into this by investigating the effects using thresholds of 5 and 15 per cent in at least two of the three subjects. By using the 5 per cent threshold I define about 2 per cent of the observations to be of low-achieving pupils, while this number increases to about 7 per cent when using the 15 per cent threshold. In both the 5 per cent and the 15 per cent approaches, when investigating the effects among all pupils combined and separately based on subject and gender, in the Appendix Tables E.1 and E.4, respectively, the effects remain very small, shifting around zero, and none of them are significantly different from zero. The same is the case when dividing the pupils based on parental education, as a proxy for socioeconomic background, in the Appendix Tables E.2 and E.5. When investigating the effects based on pupils' own position in the achievement distribution, presented in the Appendix Tables E.3 and E.6, I find some small significant estimates in both directions for pupils between the 20th and the 40th percentile and between the 40th and the 60th percentile in the achievement distribution, with negative effects for pupils in the first of these groups in all subjects and maths of 5.7 and 7.2 per cent of a standard deviation, respectively, when using the 5 per cent threshold, and a significant negative effect of 3.9 per cent of a standard deviation when using the 15 per cent threshold in all subjects combined. For the second group mentioned, pupils between the 40th and the 60th percentile, I find small positive estimates of 3.9 and 5.2 per cent of a standard deviation in all subjects and reading, respectively, when using the 15 per cent threshold.

5.5.2. Excluding pupils who are exempt from participating in the national tests

One could argue that some pupils who are exempt from participating in the national tests may not be low-achievers; they are instead pupils who are unable to cope with the testing situation (i.e. some pupils with autism, ADHD, etc.), and thus defining these pupils as low-achievers (by assigning zero as their test scores) might not be the best solution. In

Figure E.1, I show that there is in fact a close to normal distribution of test scores among pupils that are exempt from participating (and thus SEN pupils) in subjects other than the one they were exempt from. However, the number of pupils in the graph for 5th grade test scores is only 136 in reading and 474 in maths, while the corresponding numbers are 1,365 and 1,133 pupils in the graph for 9th grade test scores, respectively. In order to investigate if these pupils are driving the results in any direction, I include robustness tests where I exclude these pupils entirely from the sample. Estimated effects from these regressions are presented in the Appendix Tables E.7, E.8 and E.9, and are similar to estimated effects presented previously using the full sample.

5.5.3. Defining pupils who are exempt from participating in the national tests as low-achievers

With regards to the inclusive education literature, where the focus is SEN pupils, I also include a robustness test where I define pupils who are exempt from participating in the national tests in at least one of the three subjects in 5th grade as low-achievers. In the data, I do not observe SEN pupils directly, but in order to be exempt from participating, the pupil has to receive special education, ref. section 3 and Appendix A, making exempt pupils my best definition of SEN pupils. However, because each individual pupil actually has to be exempted, and both the pupil and the parents have to agree that the pupils should be exempt, many SEN pupils actually take the tests, so there is a weakness with this definition. The estimated effects when using this definition of low-achieving pupils are presented in the Appendix Tables E.10, E.11 and E.12. The estimated effects presented here are in line with previous findings; there are some significant effects scattered around, but none big enough to be talking about any "real" effects.

5.5.4. Pupils attending schools with few low-achievers

As previously mentioned, my identifying assumption is based on the fact that pupils within a given school are randomly assigned to classes, and this also applies for the low-achievers. However, the distribution of low-achievers across schools could be less random, making some schools have a greater population of low-achieving pupils, and thereby within certain schools, there might be a higher possibility of being placed into classes with at least one low-achieving peer. By reducing the sample to only pupils attending schools with few (less than four) low-achievers within the school, it is easier to justify the assumption about the low-achievers truly being randomly assigned, and I am thereby able to investigate if this changes the effects. It does not. There are still some

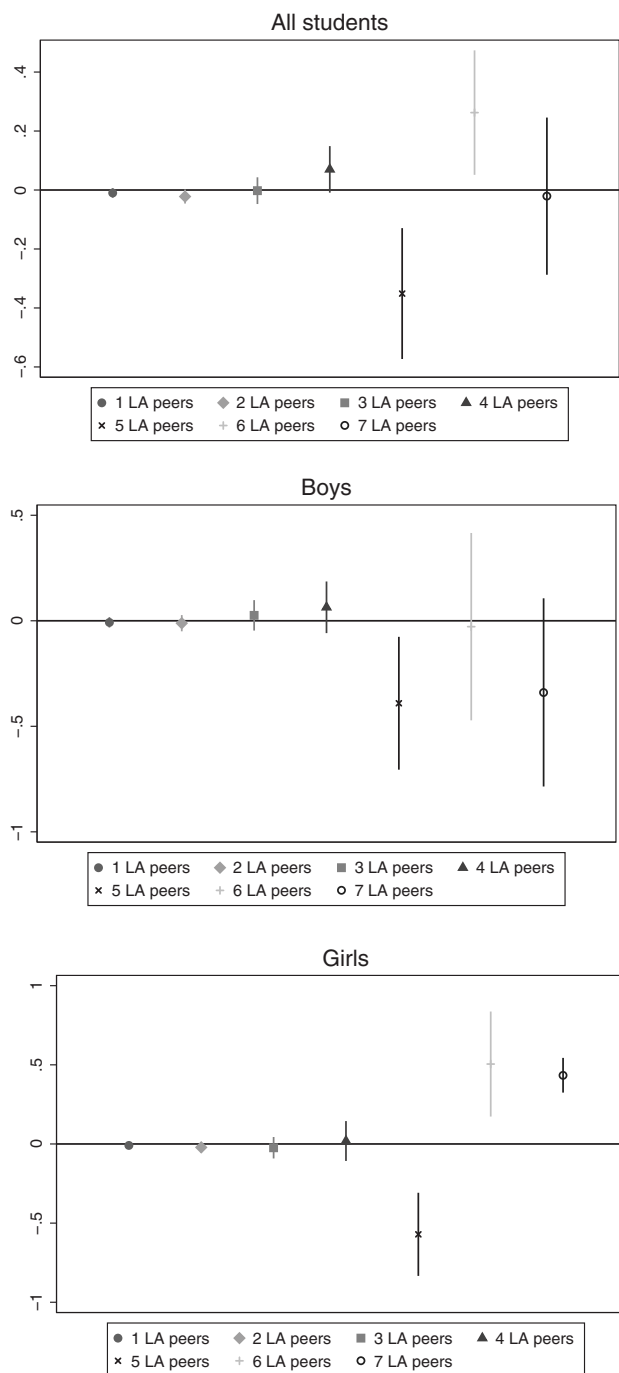


Fig. 4. Estimated coefficients based on number of low-achieving peers. Notes: Estimated coefficients for 1, 2, 3, 4, 5, 6 and 7 low-achieving peers in the classroom. First graph shows the estimated effects for all pupils combined, second and third graphs show the estimated effects for boys and girls, respectively.

small estimates scattered around, reported in the Appendix Tables E.13, E.14 and E.15, like in the other analysis, but there are no sizeable significant effects.

5.5.5. Using 8th grade test scores

As mentioned, the data also contains information about pupils' test scores in 8th grade, and by using these I am able to investigate whether there are any effects at a very early stage of the exposure to at least one low-achiever in the classroom (the autumn after the pupils have started lower secondary). Estimated effects are presented in the Appendix Tables E.16, E.17 and E.18. Since the data also contains information about pupils' test scores in English, I am able to investigate the effects in all three subjects when using the 8th grade test scores. As in the other analysis, the estimated effects remain small, shifting around zero, and most of them are not significantly different from zero.³¹

5.5.6. Excluding immigrant pupils

Another concern is the inclusion of pupils receiving special language training as a part of the low-achievers. In this last set of robustness tests, I deal with this by excluding immigrant pupils entirely from the sample and looking only at the effects of non-immigrant low-achievers on non-immigrant pupils. Estimated coefficients are presented in the Appendix Tables E.19, E.20 and E.21. The zero effects found in the previous analyses even hold in these robustness tests.

6. Discussion and interpretation

Estimated effects presented in section 5 show that there is no clear pattern that low-achieving pupils in the inclusive Norwegian compulsory education affect their peers. This means that in the present study there is no evidence for the Bad Apple model of peer effects (Hoxby and Weingarth, 2005; Lazear, 2001). This is also consistent when using the share of low-achievers within the school, when focusing on more than one low-achieving pupil in the classroom and when looking at the effects separately based on gender of both the pupil and the low-achiever combined and separately in the specification analysis. A set of robustness checks confirms that the results are not driven by the 10 per cent threshold, that pupils exempt from participating are included with zero as their test score in the subjects they are exempt from, immigrant pupils, or the fact that pupils might not be randomly assigned to classes with low-achievers in schools with a large number of low-achieving pupils.

These results are not in line with most findings in the peer effect literature, mentioned previously, reporting evidence that pupils are negatively affected by disadvantaged or low-achieving peers. In one study using a similar definition of low-achieving peers in the English secondary school, Lavy et al. (2012b), found evidence that these pupils had a significant and sizable negative effect on their peers. The findings in their study are consistent with results found in studies investigating the effect of disadvantaged and disruptive peers (Carrell et al., 2016; Carrell and Hoekstra, 2010; Imberman et al., 2012). When investigating the effects in Texas public schools, Hoxby (2000) found evidence that pupils are positively affected by the achievement level of their peers, indicating that having low-achieving peers might reduce pupils' test scores. In a later study, however, Hoxby and Weingarth (2005) found similar evidence as this study with regards to the Bad Apple model, with little evidence for the Linear-in-Means, the Shining Light and the Bad Apple models when investigating peer effects in Wake County, North Carolina, as explained in detail in their study.³²

³¹ The largest significant effect in this case is a negative effect of 5.6 per cent of a standard deviation for pupils below the 20 percentile in maths, which corresponds to 0.4 lower test scores of the exposed pupils in this group with a mean test score of 16.5 on a scale ranging from 0 to 58.

³² It is worth mentioning that even those studies that do find statistically significant effects tend to find relatively small effects (Gibbons and Telhaj, 2008).

The estimated effects in this study are more in line with studies from the inclusive education literature, reporting no differences between pupils they define as "regular" pupils in inclusive vs non-inclusive settings (Huber et al., 2001; Kalambouka et al., 2007; Ruijs, 2017; Ruijs et al., 2010). In their study, Ruijs and Peetsma (2009) highlight the need for research on different groups of pupils in order to better understand the effects, which I do with regard to several aspects such as gender, socioeconomic status and position in the achievement distribution, without this changing the results.

One explanation for different findings in Norway could be the Norwegian education system's desire to be inclusive. This is done through adapted education, meaning that all pupils, even if they do not have sufficient benefits from standard education, should experience affiliation with a class and participate in the school community when attending a Norwegian school. An OECD paper from 2009 shows that pupils in Norwegian schools report low levels of bullying and that they generally enjoy going to school. It also states that "Norwegian schools are quite successful in creating a safe and inclusive social environment at school" (Boarini, 2009, p. 13). In addition, in a study by Aizer (2008), the author reports finding that resources can help overcome negative peer effects. Viewed in conjunction with the extensive public funding in the educational sector in Norway, this could explain why I do not find any evidence of negative peer effects from the presence of at least one low-achieving pupil in the classroom (Boarini, 2009; OECD, 2017).

7. Conclusion

This paper studies the effects of having at least one low-achieving peer in lower secondary school on a pupil's achievement in the national tests. In the empirical strategy, I exploit the fact that many Norwegian school children have to change schools and thereby class when starting lower secondary, and that they are tested in the same subjects both before and after this change (both in primary and in lower secondary school) in a fixed effect approach. Thanks to massive registry data collected and maintained by Statistics Norway, I have information about pupils' test scores in 5th, 8th and 9th grade, individual characteristic information and information about pupils' parents, which I apply in this study.

I do not find any impact from the presence of at least one low-achieving pupil on his or her peers. When dividing the sample into subsamples based on subject, gender, socioeconomic status or pupils' own position in the achievement distribution, the results remain similar; neither positive nor negative effects on pupils' achievement. The results even hold when using a range of specification and robustness analyses. These findings contradict findings in the massive peer effect literature, but are more in line with the smaller inclusive education literature. These two areas of interest in the educational literature are closely related, since countries focusing on inclusive education have to spend extra resources in a normal class setting precisely because of potential peer effects and because all pupils deserve to be taken care of. With this in mind, one possible explanation for not finding any effect of low-achieving peers in this study might be precisely because Norwegian schools strive to be inclusive, and because of the extra resources put in to the educational sector as a result of this. However, in order to draw conclusions regarding the mechanisms behind the differences in educational settings more research is needed.

This study emphasises the fact that peer effects might play out very differently in different settings, and that there are a lot of considerations when investigating the effects of low-achieving peers (Angrist, 2014; Feld and Zölitz, 2017). Little is still known with regards to the mechanisms behind peer effects, and how different groups of pupils are affected by different peers. In addition, more research is necessary in order to determine the effect of inclusive education.

Table A.1
Arguments for including pupils who are exempt from participating in the national tests in 5th grade as low-achievers

<i>Panel A. Both subjects combined</i>	National tests in 5th grade	
	Exempt	Participated
Standardised Test Scores 9th grade	-1.847	0.056
Exempt 9th grade	0.287	0.005
Immigrants	0.479	0.282
Observations	2 695	88 250
	National tests in 5th grade	
<i>Panel B. Reading</i>	Exempt	Participated
Standardised Test Scores 9th grade	-1.832	0.071
Exempt 9th grade	0.270	0.005
Immigrants	0.488	0.281
Observations	1 701	43 821
	National tests in 5th grade	
<i>Panel C. Maths</i>	Exempt	Participated
Standardised Test Scores 9th grade	-1.874	0.042
Exempt 9th grade	0.316	0.006
Immigrants	0.463	0.283
Observations	994	44 429

Notes: Mean statistics for pupils who were exempt and pupils who participated in the national tests in 5th grade. The first column contains pupils who were exempt from participating, while the second column contains pupils who participated in the national tests in 5th grade. The first row shows pupils' standardised test scores in the national tests in 9th grade, the second row shows how likely it is for pupils to also be exempt from participating in 9th grade and the third row shows how likely the pupils are to be immigrants. The "Observations" row reports the number of observations within the different categories. Panel A shows the statistics in both subjects combined, while Panels B and C show the statistics in reading and maths, respectively.

Appendix A. Participation status in the national tests³³

The national tests in 5th, 8th and 9th grade are mandatory and all pupils have the right to participate in these tests. However, pupils who require special education or special language training may be exempt from participating if it is clear that the result from the tests will not affect their future learning, ref. section 2 and below with regards to who receives special education. In each individual case, it is the head of the school in cooperation with the pupil's teacher that determines who is granted an exemption. Exempt pupils or their parents can decide whether the pupil takes the tests even where the head of the school has deemed they should be exempt.³⁴

*Pupils who require special education and special language training*³⁵. As mentioned in section 2, the crucial factor for receiving special education or special language training is that the pupil is unable to follow standard education, and this indicates that the pupil is considered to be a low-achiever. In Appendix Table A.1, I show that pupils exempt from participating in the national tests in 5th grade have lower test scores in the national tests in 9th grade. They are more likely to be exempt from participating in the national tests in 9th grade as well, and are more likely to be immigrants. Because pupils who are exempt from participating in the national tests are the pupils who do not have sufficient benefits from a standard educational setting (i.e. they have difficulties following the standard curriculum), and who I would expect to make up the lower tail of the achievement distribution if they were included, I include these pupils when making the national subject- and cohort-specific achievement distribution, used to determine if a pupil is a low-achiever or not.³⁶

³³ The Norwegian Directorate for Education and Training (2018a)

³⁴ It is possible to only be exempt from one subject.

³⁵ The Norwegian Directorate for Education and Training (2016, 2018b)

³⁶ They are included with zero as their test score.

Appendix B. Distribution of pupils in schools with several low-achieving pupils

In Appendix Figure B.1, I show the distribution of pupils in classes with different numbers of low-achieving peers depending on how many low-achievers there are within the school. For example, in schools with three low-achievers within the school (the graph labelled “3” in Appendix Figure B.1), about 45 per cent of the pupils attend a class without any low-achievers, 35 per cent attend a class with one low-achieving peer and 20 per cent attend classes with two low-achieving peers. The graphs show that even in schools with several low-achievers, there are pupils attending classes without any low-achieving peers.³⁷

³⁷ Appendix Figure B.2 shows the same distribution with regards to SEN pupils.

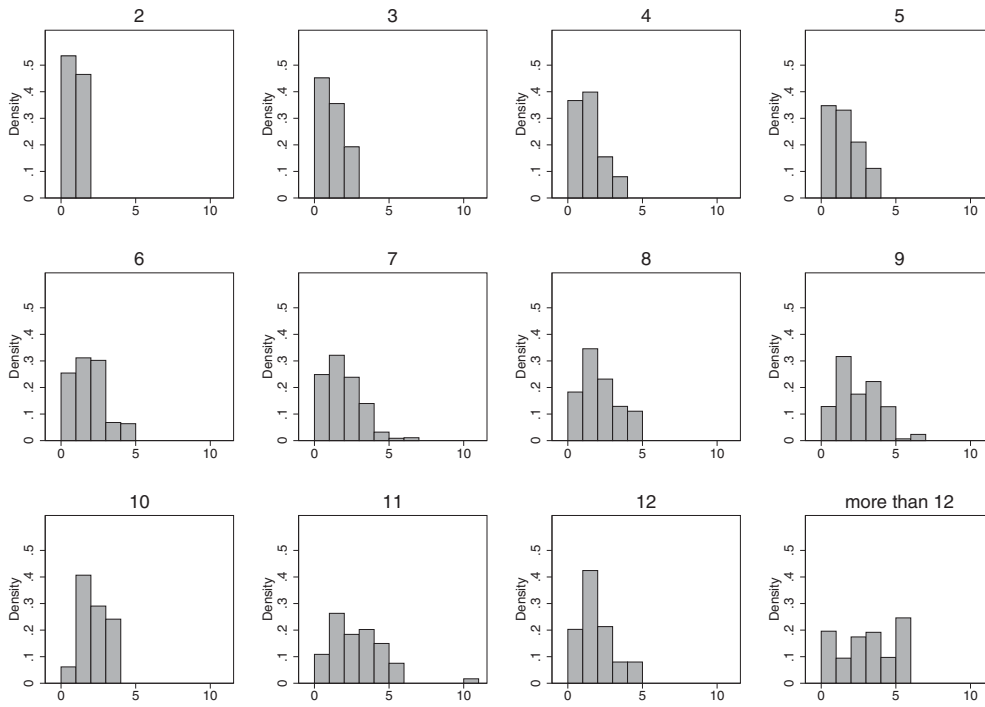


Fig. B.1. Distribution of low-achievers within schools with different numbers of low achieving pupils. Notes: Distribution of low-achieving pupils in classrooms with different numbers of low-achieving pupils within schools with 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12 and more than 12 low-achieving pupils within the school in the different graphs.

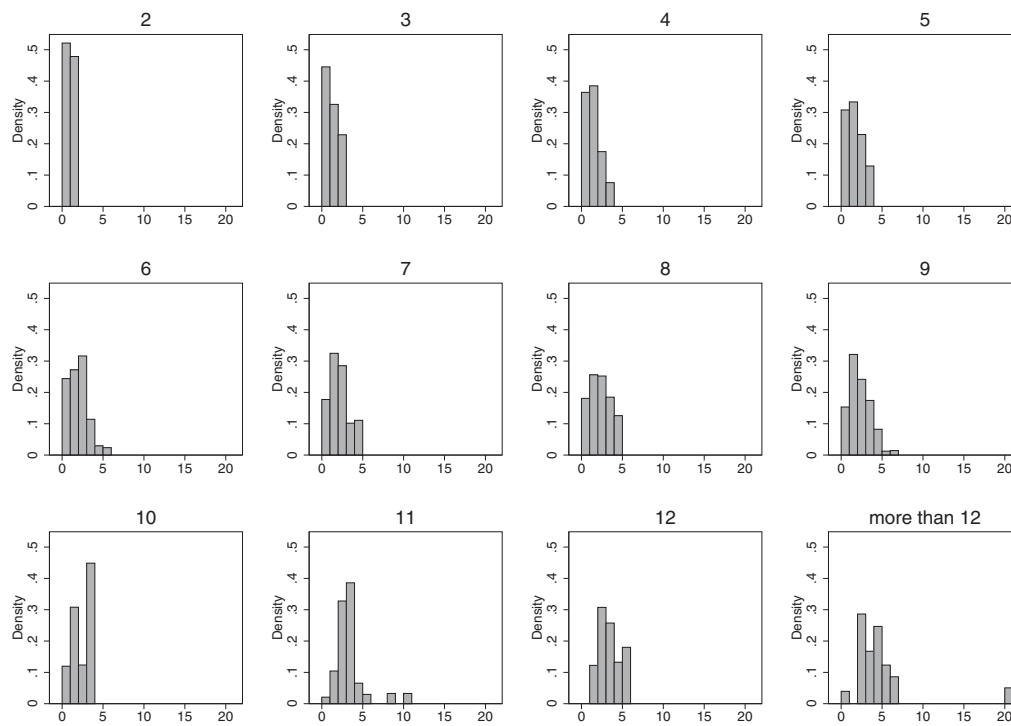


Fig. B.2. Distribution of SEN pupils within schools with different numbers of SEN pupils. Notes: Distribution of SEN pupils in classrooms with different numbers of SEN pupils within schools with 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12 and more than 12 SEN pupils within the school in the different graphs.

Appendix C. Subsample by subject and parental income

In Table C.1, I examine different groups of pupils based on parental income, as a proxy for socioeconomic status. In this analysis I find similar results as previously presented; no sizeable significant estimates indicating that any groups of pupils are affected by the presence of at least one low-achieving peer in the classroom.

Table C.1
Subsample analysis by subject and parental income

	Parental income				
	Below 20th percentile	20th to 40th percentile	40th to 60th percentile	60th to 80th percentile	Above 80th percentile
<i>Panel A. Both subjects</i>					
Having "low-achieving" peer(s)	0.012 (0.023)	-0.026 (0.019)	-0.014 (0.017)	-0.009 (0.015)	-0.013 (0.016)
<i>Panel B. Reading</i>					
Having "low-achieving" peer(s)	0.003 (0.035)	-0.037 (0.028)	0.004 (0.025)	-0.002 (0.022)	-0.024 (0.023)
<i>Panel C. Maths</i>					
Having "low-achieving" peer(s)	0.024 (0.030)	-0.012 (0.024)	-0.031 (0.024)	-0.016 (0.020)	-0.003 (0.020)

Notes: Fixed effects regressions for subsamples of pupils in both subjects combined and separately based on subject and parental income, with standard errors clustered on class in 5th grade and school in 9th grade in parentheses. Significance levels are indicated as follows: + significant at 10 per cent; * significant at 5 per cent; ** significant at 1 per cent. Dependent variable is standardised test scores in national tests in 9th grade, and the coefficient of interest is the coefficient on the dummy for having low-achieving peer(s) in 9th grade. Coefficients are from separate regressions, based on model four (4) in Table 3, where I include fixed effects for the interaction between class in 5th grade and school in 9th grade, and controls for individual and parental characteristics. Parental income is specified in columns 1 to 5, with the lowest income parents in column 1 and the highest income parents in column 5. In Panel A, I look at both subjects combined, and in Panels B and C I look only at test scores in reading and maths, respectively. The percentiles are within each panel, meaning that pupils below the 20th percentile in Panel A are not necessarily below the 20th percentile in Panels B and/or C.

Appendix D

Table D.1
Subsample analysis boys - subject and parental education

	Mother's education			Father's education		
	Not completed upper secondary school	Completed upper secondary school	University Degree	Not completed upper secondary school	Completed upper secondary school	University Degree
<i>Panel A. Both subjects</i>						
Having "low-achieving" peer(s)	-0.012 (0.028)	-0.006 (0.020)	-0.019 (0.017)	-0.047* (0.026)	-0.021 (0.017)	0.004 (0.019)
<i>Panel C. Reading</i>						
Having "low-achieving" peer(s)	-0.000 (0.041)	-0.016 (0.031)	-0.026 (0.025)	-0.066* (0.038)	-0.013 (0.026)	0.020 (0.029)
<i>Panel D. Maths</i>						
Having "low-achieving" peer(s)	-0.024 (0.038)	0.003 (0.026)	-0.010 (0.022)	-0.027 (0.036)	-0.028 (0.023)	-0.011 (0.025)

Notes: Fixed effects regressions for subsamples of boys in both subjects combined and separately based on subject and parental education, with standard errors clustered on class in 5th grade and school in 9th grade in parentheses. Significance levels are indicated as follows: + significant at 10 per cent; * significant at 5 per cent; ** significant at 1 per cent. Dependent variable is standardised test scores in national tests in 9th grade, and the coefficient of interest is the coefficient on the dummy for having low-achieving peer(s) in 9th grade. Coefficients are from separate regressions, based on model four (4) in Table 3, where I include fixed effects for the interaction between class in 5th grade and school in 9th grade, and controls for individual and parental characteristics. Parental education is specified in columns 1 to 6, with mother's education in columns 1 to 3 and father's education in columns 4 to 6. In Panel A, I look at both subjects combined, and in Panels B and C I look only at test scores in reading and maths, respectively.

Table D.2
Subsample analysis boys - subject and pupils' own position in the achievement distribution

	Pupils' own position in the achievement distribution				
	Below 20th percentile	20th to 40th percentile	40th to 60th percentile	60th to 80th percentile	Above 80th percentile
<i>Panel A. Both subjects</i>					
Having "low-achieving" peer(s)	-0.004 (0.045)	-0.033 (0.033)	0.033 (0.031)	-0.013 (0.028)	0.026 (0.022)
<i>Panel C. Reading</i>					
Having "low-achieving" peer(s)	0.071 (0.064)	-0.083* (0.043)	0.079* (0.047)	0.020 (0.043)	0.031 (0.032)
<i>Panel D. Maths</i>					
Having "low-achieving" peer(s)	-0.101* (0.061)	0.032 (0.052)	-0.013 (0.040)	-0.045 (0.036)	0.022 (0.030)

Notes: Fixed effects regressions for subsamples of boys in both subjects combined and separately based on subject and pupils' own position in the achievement distribution, with standard errors clustered on class in 5th grade and school in 9th grade in parentheses. Significance levels are indicated as follows: + significant at 10 per cent; * significant at 5 per cent; ** significant at 1 per cent. Dependent variable is standardised test scores in national tests in 9th grade, and the coefficient of interest is the coefficient on the dummy for having low-achieving peer(s) in 9th grade. Coefficients are from separate regressions, based on model four (4) in Table 3, where I include fixed effects for the interaction between class in 5th grade and school in 9th grade, and controls for individual and parental characteristics. Pupils' own position in the achievement distribution is specified in columns 1 to 5, with the lowest achieving pupils in column 1 and the highest achieving pupils in column 5. In Panel A, I look at both subjects combined, and in Panels B and C I look only at test scores in reading and maths, respectively.

Table D.3
Subsample analysis girls - subject and parental education

	Mother's education			Father's education		
	Not completed upper secondary school	Completed upper secondary school	University Degree	Not completed upper secondary school	Completed upper secondary school	University Degree
<i>Panel A. Both subjects</i>						
Having "low-achieving" peer(s)	0.005 (0.025)	-0.000 (0.018)	-0.024 (0.016)	0.031 (0.024)	-0.019 (0.016)	-0.047* (0.018)
<i>Panel C. Reading</i>						
Having "low-achieving" peer(s)	0.030 (0.038)	-0.008 (0.025)	-0.033 (0.022)	0.057 (0.036)	-0.023 (0.023)	-0.040 (0.026)
<i>Panel D. Maths</i>						
Having "low-achieving" peer(s)	-0.021 (0.033)	0.009 (0.026)	-0.018 (0.023)	0.006 (0.032)	-0.017 (0.021)	-0.055* (0.026)

Notes: Fixed effects regressions for subsamples of girls in both subjects combined and separately based on subject and parental education, with standard errors clustered on class in 5th grade and school in 9th grade in parentheses. Significance levels are indicated as follows: + significant at 10 per cent; * significant at 5 per cent; ** significant at 1 per cent. Dependent variable is standardised test scores in national tests in 9th grade, and the coefficient of interest is the coefficient on the dummy for having low-achieving peer(s) in 9th grade. Coefficients are from separate regressions, based on model four (4) in Table 3, where I include fixed effects for the interaction between class in 5th grade and school in 9th grade, and controls for individual and parental characteristics. Parental education is specified in columns 1 to 6, with mother's education in columns 1 to 3 and father's education in columns 4 to 6. In Panel A, I look at both subjects combined, and in Panels B and C I look only at test scores in reading and maths, respectively.

Table D.4
Subsample analysis girls - subject and pupils' own position in the achievement distribution

	Pupils' own position in the achievement distribution				
	Below 20th percentile	20th to 40th percentile	40th to 60th percentile	60th to 80th percentile	Above 80th percentile
<i>Panel A. Both subjects</i>					
Having "low-achieving" peer(s)	-0.004 (0.049)	-0.028 (0.030)	0.007 (0.026)	-0.010 (0.024)	-0.032 (0.022)
<i>Panel C. Reading</i>					
Having "low-achieving" peer(s)	0.041 (0.085)	-0.046 (0.044)	0.012 (0.039)	0.021 (0.033)	-0.037 (0.027)
<i>Panel D. Maths</i>					
Having "low-achieving" peer(s)	-0.035 (0.059)	-0.009 (0.042)	-0.009 (0.036)	-0.046 (0.035)	-0.027 (0.036)

Notes: Fixed effects regressions for subsamples of girls in both subjects combined and separately based on subject and pupils' own position in the achievement distribution, with standard errors clustered on class in 5th grade and school in 9th grade in parentheses. Significance levels are indicated as follows: + significant at 10 per cent; * significant at 5 per cent; ** significant at 1 per cent. Dependent variable is standardised test scores in national tests in 9th grade, and the coefficient of interest is the coefficient on the dummy for having low-achieving peer(s) in 9th grade. Coefficients are from separate regressions, based on model four (4) in Table 3, where I include fixed effects for the interaction between class in 5th grade and school in 9th grade, and controls for individual and parental characteristics. Pupils' own position in the achievement distribution is specified in columns 1 to 5, with the lowest achieving pupils in column 1 and the highest achieving pupils in column 5. In Panel A, I look at both subjects combined, and in Panels B and C I look only at test scores in reading and maths, respectively.

Table D.5
Specification analysis by gender of the low-achieving pupil - subject and gender

	All pupils	Gender	
		Boys	Girls
<i>Panel A. Both subjects</i>			
Having "low-achieving" boy(s)	-0.014* (0.007)	-0.014 (0.011)	-0.007 (0.010)
Having "low-achieving" girl(s)	0.010 (0.008)	0.020* (0.012)	0.005 (0.011)
<i>Panel C. Reading</i>			
Having "low-achieving" boy(s)	-0.017* (0.010)	-0.019 (0.016)	-0.009 (0.014)
Having "low-achieving" girl(s)	0.015 (0.011)	0.026 (0.018)	0.013 (0.016)
<i>Panel D. Maths</i>			
Having "low-achieving" boy(s)	-0.011 (0.009)	-0.009 (0.014)	-0.007 (0.014)
Having "low-achieving" girl(s)	0.005 (0.010)	0.012 (0.015)	-0.004 (0.015)

Notes: Fixed effects regressions for all pupils in both subjects combined and separately based on subject and gender, with standard errors clustered on class in 5th grade and school in 9th grade in parentheses. Significance levels are indicated as follows: + significant at 10 per cent; * significant at 5 per cent; ** significant at 1 per cent. Dependent variable is standardised test scores in national tests in 9th grade, and the coefficients of interest are the coefficients on the dummy for having low-achieving boy(s) and low-achieving girl(s) in 9th grade. Coefficients in different panels and columns are from separate regressions, based on equation (2), where I include fixed effects for the interaction between class in 5th grade and school in 9th grade, and controls for individual and parental characteristics. The first column represents all pupils combined, the second represents boys and the third column represents girls. In Panel A, I look at both subjects combined, and in Panels B and C I look only at test scores in reading and maths, respectively.

Table D.6
Specification analysis by gender of the low-achieving pupil - subject and parental education

	Mother's education			Father's education		
	Not completed upper secondary school	Completed upper secondary school	University Degree	Not completed upper secondary school	Completed upper secondary school	University Degree
<i>Panel A. Both subjects</i>						
Having "low-achieving" boy(s)	-0.021 (0.017)	-0.019 (0.013)	-0.014 (0.011)	-0.006 (0.016)	-0.020* (0.011)	-0.012 (0.012)
Having "low-achieving" girl(s)	0.012 (0.017)	0.028* (0.014)	0.000 (0.012)	-0.002 (0.018)	0.014 (0.012)	-0.004 (0.014)
<i>Panel C. Reading</i>						
Having "low-achieving" boy(s)	-0.020 (0.025)	-0.028 (0.019)	-0.021 (0.016)	0.006 (0.024)	-0.033* (0.017)	-0.016 (0.018)
Having "low-achieving" girl(s)	0.013 (0.025)	0.029 (0.021)	-0.000 (0.017)	-0.016 (0.026)	0.027 (0.018)	0.013 (0.020)
<i>Panel D. Maths</i>						
Having "low-achieving" boy(s)	-0.023 (0.022)	-0.008 (0.017)	-0.006 (0.015)	-0.017 (0.020)	-0.007 (0.015)	-0.008 (0.017)
Having "low-achieving" girl(s)	0.013 (0.023)	0.024 (0.019)	0.001 (0.016)	0.014 (0.024)	0.001 (0.016)	-0.024 (0.019)

Notes: Fixed effects regressions for subsamples of pupils in both subjects combined and separately based on subject and parental education, with standard errors clustered on class in 5th grade and school in 9th grade in parentheses. Significance levels are indicated as follows: + significant at 10 per cent; * significant at 5 per cent; ** significant at 1 per cent. Dependent variable is standardised test scores in national tests in 9th grade, and the coefficients of interest are the coefficients on the dummy for having low-achieving boy(s) and low-achieving girl(s) in 9th grade. Coefficients in different panels and columns are from separate regressions, based on equation (2), where I include fixed effects for the interaction between class in 5th grade and school in 9th grade, and controls for individual and parental characteristics. Parental education is specified in columns 1 to 6, with mother's education in columns 1 to 3 and father's education in columns 4 to 6. In Panel A, I look at both subjects combined, and in Panels B and C I look only at test scores in reading and maths, respectively.

Table D.7
Specification analysis by gender of the low-achieving pupil - subject and pupils' own position in the achievement distribution

	Pupils' own position in the achievement distribution				
	Below 20th percentile	20th to 40th percentile	40th to 60th percentile	60th to 80th percentile	Above 80th percentile
<i>Panel A. Both subjects</i>					
Having "low-achieving" boy(s)	-0.016 (0.026)	-0.024 (0.020)	-0.019 (0.018)	-0.031 ⁺ (0.016)	0.012 (0.014)
Having "low-achieving" girl(s)	0.017 (0.030)	-0.005 (0.021)	0.043* (0.020)	0.020 (0.017)	-0.007 (0.016)
<i>Panel C. Reading</i>					
Having "low-achieving" boy(s)	-0.002 (0.041)	-0.041 (0.028)	-0.014 (0.027)	-0.018 (0.023)	-0.005 (0.019)
Having "low-achieving" girl(s)	0.071 (0.046)	-0.032 (0.030)	0.069* (0.029)	0.041 (0.025)	-0.009 (0.021)
<i>Panel D. Maths</i>					
Having "low-achieving" boy(s)	-0.029 (0.033)	-0.004 (0.027)	-0.024 (0.024)	-0.049* (0.022)	0.026 (0.020)
Having "low-achieving" girl(s)	-0.030 (0.038)	0.023 (0.029)	0.016 (0.028)	-0.001 (0.023)	-0.006 (0.023)

Notes: Fixed effects regressions for subsamples of pupils in both subjects combined and separately based on subject and pupils' own position in the achievement distribution, with standard errors clustered on class in 5th grade and school in 9th grade in parentheses. Significance levels are indicated as follows: + significant at 10 per cent; * significant at 5 per cent; ** significant at 1 per cent. Dependent variable is standardised test scores in national tests in 9th grade, and the coefficients of interest are the coefficients on the dummy for having low-achieving boy(s) and low-achieving girl(s) in 9th grade. Coefficients in different panels and columns are from separate regressions, based on equation (2), where I include fixed effects for the interaction between class in 5th grade and school in 9th grade, and controls for individual and parental characteristics. Pupils' own position in the achievement distribution is specified in columns 1 to 5, with the lowest achieving pupils in column 1 and the highest achieving pupils in column 5. In Panel A, I look at both subjects combined, and in Panels B and C I look only at test scores in reading and maths, respectively.

Table D.8
Subsample analysis of boys by gender of the low-achieving pupil - subject and parental education

	Mother's education			Father's education		
	Not completed upper secondary school	Completed upper secondary school	University Degree	Not completed upper secondary school	Completed upper secondary school	University Degree
<i>Panel A. Both subjects</i>						
Having "low-achieving" boy(s)	0.012 (0.027)	-0.009 (0.021)	-0.023 (0.018)	-0.030 (0.026)	-0.025 (0.018)	0.012 (0.020)
Having "low-achieving" girl(s)	-0.004 (0.029)	0.011 (0.024)	0.015 (0.019)	0.014 (0.030)	0.007 (0.019)	0.014 (0.022)
<i>Panel C. Reading</i>						
Having "low-achieving" boy(s)	0.024 (0.040)	-0.020 (0.032)	-0.031 (0.026)	-0.044 (0.039)	-0.030 (0.027)	0.019 (0.029)
Having "low-achieving" girl(s)	-0.020 (0.043)	0.015 (0.036)	0.018 (0.029)	-0.004 (0.044)	0.024 (0.028)	0.029 (0.033)
<i>Panel D. Maths</i>						
Having "low-achieving" boy(s)	-0.004 (0.036)	0.002 (0.028)	-0.013 (0.023)	-0.012 (0.035)	-0.022 (0.024)	0.006 (0.027)
Having "low-achieving" girl(s)	0.012 (0.038)	0.004 (0.031)	0.012 (0.026)	0.033 (0.041)	-0.011 (0.024)	0.000 (0.030)

Notes: Fixed effects regressions for subsamples of boys in both subjects combined and separately based on subject and parental education, with standard errors clustered on class in 5th grade and school in 9th grade in parentheses. Significance levels are indicated as follows: + significant at 10 per cent; * significant at 5 per cent; ** significant at 1 per cent. Dependent variable is standardised test scores in national tests in 9th grade, and the coefficients of interest are the coefficients on the dummy for having low-achieving boy(s) and low-achieving girl(s) in 9th grade. Coefficients in different panels and columns are from separate regressions, based on equation (2), where I include fixed effects for the interaction between class in 5th grade and school in 9th grade, and controls for individual and parental characteristics. Parental education is specified in columns 1 to 6, with mother's education in columns 1 to 3 and father's education in columns 4 to 6. In Panel A, I look at both subjects combined, and in Panels B and C I look only at test scores in reading and maths, respectively.

Table D.9
Subsample analysis of boys by gender of the low-achieving pupil - subject and pupils' own position in the achievement distribution

	Pupils' own position in the achievement distribution				
	Below 20th percentile	20th to 40th percentile	40th to 60th percentile	60th to 80th percentile	Above 80th percentile
<i>Panel A. Both subjects</i>					
Having "low-achieving" boy(s)	-0.001 (0.043)	-0.029 (0.033)	-0.008 (0.032)	-0.044 (0.030)	0.029 (0.024)
Having "low-achieving" girl(s)	0.044 (0.051)	0.008 (0.037)	0.081* (0.035)	0.039 (0.032)	0.003 (0.026)
<i>Panel C. Reading</i>					
Having "low-achieving" boy(s)	0.035 (0.062)	-0.061 (0.043)	-0.028 (0.048)	0.024 (0.046)	0.007 (0.034)
Having "low-achieving" girl(s)	0.104 (0.073)	-0.032 (0.049)	0.136** (0.049)	0.020 (0.051)	0.010 (0.040)
<i>Panel D. Maths</i>					
Having "low-achieving" boy(s)	-0.051 (0.059)	0.014 (0.052)	0.016 (0.041)	-0.109** (0.039)	0.045 (0.033)
Having "low-achieving" girl(s)	-0.032 (0.069)	0.059 (0.055)	0.018 (0.049)	0.059 (0.038)	-0.001 (0.035)

Notes: Fixed effects regressions for subsamples of boys in both subjects combined and separately based on subject and pupils' own position in the achievement distribution, with standard errors clustered on class in 5th grade and school in 9th grade in parentheses. Significance levels are indicated as follows: + significant at 10 per cent; * significant at 5 per cent; ** significant at 1 per cent. Dependent variable is standardised test scores in national tests in 9th grade, and the coefficients of interest are the coefficients on the dummy for having low-achieving boy(s) and low-achieving girl(s) in 9th grade. Coefficients in different panels and columns are from separate regressions, based on equation (2), where I include fixed effects for the interaction between class in 5th grade and school in 9th grade, and controls for individual and parental characteristics. Pupils' own position in the achievement distribution is specified in columns 1 to 5, with the lowest achieving pupils in column 1 and the highest achieving pupils in column 5. In Panel A, I look at both subjects combined, and in Panels B and C I look only at test scores in reading and maths, respectively.

Table D.10
Subsample analysis of girls by gender of the low-achieving pupil - subject and parental education

	Mother's education			Father's education		
	Not completed upper secondary school	Completed upper secondary school	University Degree	Not completed upper secondary school	Completed upper secondary school	University Degree
<i>Panel A. Both subjects</i>						
Having "low-achieving" boy(s)	-0.024 (0.026)	-0.019 (0.018)	-0.013 (0.016)	0.031 (0.025)	-0.027* (0.016)	-0.033* (0.020)
Having "low-achieving" girl(s)	0.029 (0.027)	0.041* (0.020)	-0.013 (0.018)	0.009 (0.027)	0.020 (0.018)	-0.014 (0.023)
<i>Panel C. Reading</i>						
Having "low-achieving" boy(s)	-0.026 (0.039)	-0.027 (0.026)	-0.015 (0.023)	0.070+ (0.037)	-0.036 (0.024)	-0.033 (0.028)
Having "low-achieving" girl(s)	0.071+ (0.041)	0.028 (0.028)	-0.007 (0.025)	0.002 (0.041)	0.023 (0.025)	-0.003 (0.032)
<i>Panel D. Maths</i>						
Having "low-achieving" boy(s)	-0.023 (0.034)	-0.009 (0.026)	-0.012 (0.023)	-0.009 (0.033)	-0.018 (0.022)	-0.035 (0.028)
Having "low-achieving" girl(s)	-0.014 (0.036)	0.055* (0.028)	-0.019 (0.026)	0.016 (0.035)	0.018 (0.025)	-0.028 (0.033)

Notes: Fixed effects regressions for subsamples of girls in both subjects combined and separately based on subject and parental education, with standard errors clustered on class in 5th grade and school in 9th grade in parentheses. Significance levels are indicated as follows: + significant at 10 per cent; * significant at 5 per cent; ** significant at 1 per cent. Dependent variable is standardised test scores in national tests in 9th grade, and the coefficients of interest are the coefficients on the dummy for having low-achieving boy(s) and low-achieving girl(s) in 9th grade. Coefficients in different panels and columns are from separate regressions, based on equation (2), where I include fixed effects for the interaction between class in 5th grade and school in 9th grade, and controls for individual and parental characteristics. Parental education is specified in columns 1 to 6, with mother's education in columns 1 to 3 and father's education in columns 4 to 6. In Panel A, I look at both subjects combined, and in Panels B and C I look only at test scores in reading and maths, respectively.

Table D.11
Subsample analysis of girls by gender of the low-achieving pupil - subject and pupils' own position in the achievement distribution

	Pupils' own position in the achievement distribution				
	Below 20th percentile	20th to 40th percentile	40th to 60th percentile	60th to 80th percentile	Above 80th percentile
<i>Panel A. Both subjects</i>					
Having "low-achieving" boy(s)	-0.023 (0.045)	-0.016 (0.034)	-0.024 (0.029)	-0.008 (0.025)	0.001 (0.022)
Having "low-achieving" girl(s)	-0.002 (0.053)	-0.021 (0.035)	0.044 (0.030)	0.010 (0.027)	-0.035 (0.025)
<i>Panel C. Reading</i>					
Having "low-achieving" boy(s)	-0.106 (0.081)	-0.034 (0.050)	0.003 (0.043)	0.007 (0.034)	-0.017 (0.029)
Having "low-achieving" girl(s)	0.173 ⁺ (0.089)	-0.030 (0.052)	0.032 (0.040)	0.062 ⁺ (0.038)	-0.045 (0.031)
<i>Panel D. Maths</i>					
Having "low-achieving" boy(s)	0.023 (0.052)	0.000 (0.045)	-0.058 (0.039)	-0.031 (0.036)	0.023 (0.035)
Having "low-achieving" girl(s)	-0.111 ⁺ (0.064)	-0.011 (0.048)	0.045 (0.043)	-0.051 (0.040)	-0.023 (0.042)

Notes: Fixed effects regressions for subsamples of girls in both subjects combined and separately based on subject and pupils' own position in the achievement distribution, with standard errors clustered on class in 5th grade and school in 9th grade in parentheses. Significance levels are indicated as follows: + significant at 10 per cent; * significant at 5 per cent; ** significant at 1 per cent. Dependent variable is standardised test scores in national tests in 9th grade, and the coefficients of interest are the coefficients on the dummy for having low-achieving boy(s) and low-achieving girl(s) in 9th grade. Coefficients in different panels and columns are from separate regressions, based on equation (2), where I include fixed effects for the interaction between class in 5th grade and school in 9th grade, and controls for individual and parental characteristics. Pupils' own position in the achievement distribution is specified in columns 1 to 5, with the lowest achieving pupils in column 1 and the highest achieving pupils in column 5. In Panel A, I look at both subjects combined, and in Panels B and C I look only at test scores in reading and maths, respectively.

Appendix E

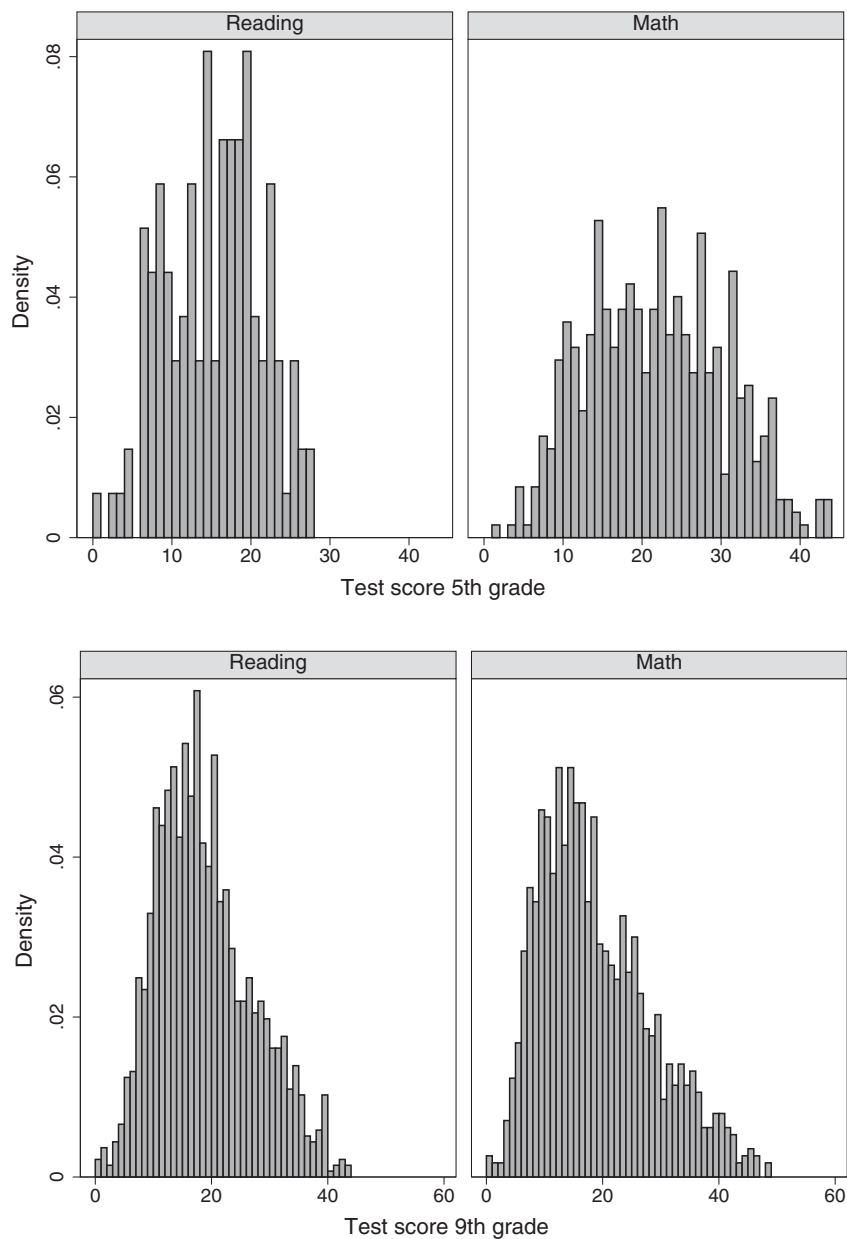


Fig. E.1. Test scores in the national tests for pupils who are exempt from participating in at least one subject. Notes: Distribution of test scores among pupils who are exempt from participating in the national test in at least one subject, for national tests they are not exempt from participating in.

Table E.1
Robustness tests using a threshold of 5 per cent - subject and gender

	All pupils	Gender		Observations
		Boys	Girls	
<i>Panel A. Both subjects</i>				
Having “low-achieving” peer(s)	-0.002 (0.008)	-0.004 (0.012)	0.002 (0.011)	89 191
<i>Panel C. Reading</i>				
Having “low-achieving” peer(s)	-0.002 (0.011)	-0.003 (0.018)	0.001 (0.016)	44 629
<i>Panel D. Maths</i>				
Having “low-achieving” peer(s)	-0.001 (0.010)	-0.004 (0.015)	0.002 (0.015)	44 562
Observations	89 191	44 840	44 351	

Notes: Fixed effects regressions for all pupils in both subjects combined and separately based on subject and gender, with standard errors clustered on class in 5th grade and school in 9th grade in parentheses. Significance levels are indicated as follows: + significant at 10 per cent; * significant at 5 per cent; ** significant at 1 per cent. Dependent variable is standardised test scores in national tests in 9th grade, and the coefficient of interest is the coefficient on the dummy for having low-achieving peer(s) in 9th grade. Coefficients are from separate regressions, based on model four (4) in Table 3, where I include fixed effects for the interaction between class in 5th grade and school in 9th grade, and controls for individual and parental characteristics. The first column represents all pupils combined, the second represents boys and the third column represents girls. In Panel A, I look at both subjects combined, and in Panels B and C I look only at test scores in reading and maths, respectively. The “Observations” column and row include the number of observations within the different categories.

Table E.2
Robustness test using a threshold of 5 per cent - subject and parental education

	Mother’s education			Father’s education		
	Not completed upper secondary school	Completed upper secondary school	University Degree	Not completed upper secondary school	Completed upper secondary school	University Degree
<i>Panel A. Both subjects</i>						
Having “low-achieving” peer(s)	-0.018 (0.018)	0.010 (0.014)	-0.013 (0.012)	0.002 (0.017)	-0.011 (0.012)	0.006 (0.013)
<i>Panel C. Reading</i>						
Having “low-achieving” peer(s)	-0.024 (0.026)	0.008 (0.020)	-0.022 (0.018)	0.020 (0.025)	-0.017 (0.018)	-0.001 (0.019)
<i>Panel D. Maths</i>						
Having “low-achieving” peer(s)	-0.011 (0.024)	0.012 (0.019)	-0.004 (0.016)	-0.017 (0.022)	-0.006 (0.016)	0.013 (0.018)

Notes: Fixed effects regressions for subsamples of pupils in both subjects combined and separately based on subject and parental education, with standard errors clustered on class in 5th grade and school in 9th grade in parentheses. Significance levels are indicated as follows: + significant at 10 per cent; * significant at 5 per cent; ** significant at 1 per cent. Dependent variable is standardised test scores in national tests in 9th grade, and the coefficient of interest is the coefficient on the dummy for having low-achieving peer(s) in 9th grade. Coefficients are from separate regressions, based on model four (4) in Table 3, where I include fixed effects for the interaction between class in 5th grade and school in 9th grade, and controls for individual and parental characteristics. Parental education is specified in columns 1 to 6, with mother’s education in columns 1 to 3 and father’s education in columns 4 to 6. In Panel A, I look at both subjects combined, and in Panels B and C I look only at test scores in reading and maths, respectively.

Table E.3
Robustness test using a threshold of 5 per cent - subject and pupils' own position in the achievement distribution

	Pupils' own position in the achievement distribution				
	Below 20th percentile	20th to 40th percentile	40th to 60th percentile	60th to 80th percentile	Above 80th percentile
<i>Panel A. Both subjects</i>					
Having "low-achieving" peer(s)	0.011 (0.027)	-0.057** (0.021)	0.000 (0.020)	0.023 (0.018)	0.018 (0.015)
<i>Panel C. Reading</i>					
Having "low-achieving" peer(s)	0.011 (0.041)	-0.044 (0.030)	0.014 (0.030)	0.020 (0.027)	-0.001 (0.021)
<i>Panel D. Maths</i>					
Having "low-achieving" peer(s)	0.014 (0.034)	-0.072* (0.029)	-0.012 (0.028)	0.026 (0.024)	0.034 (0.021)

Notes: Fixed effects regressions for subsamples of pupils in all subjects combined and separately based on subject and pupils' own position in the achievement distribution, with standard errors clustered on class in 5th grade and school in 9th grade in parentheses. Significance levels are indicated as follows: + significant at 10 per cent; * significant at 5 per cent; ** significant at 1 per cent. Dependent variable is standardised test scores in national tests in 9th grade, and the coefficient of interest is the coefficient on the dummy for having low-achieving peer(s) in 9th grade. Coefficients are from separate regressions, based on model four (4) in Table 3, where I include fixed effects for the interaction between class in 5th grade and school in 9th grade, and controls for individual and parental characteristics. Pupils' own position in the achievement distribution is specified in columns 1 to 5, with the lowest achieving pupils in column 1 and the highest achieving pupils in column 5. In Panel A, I look at both subjects combined, and in Panels B and C I look only at test scores in reading and maths, respectively.

Table E.4
Robustness test using a threshold of 15 per cent - subject and gender

	All pupils	Gender		Observations
		Boys	Girls	
<i>Panel A. Both subjects</i>				
Having "low-achieving" peer(s)	0.002 (0.007)	0.010 (0.012)	-0.001 (0.010)	84 834
<i>Panel C. Reading</i>				
Having "low-achieving" peer(s)	0.010 (0.011)	0.010 (0.018)	0.014 (0.015)	42 451
<i>Panel D. Maths</i>				
Having "low-achieving" peer(s)	-0.006 (0.010)	0.010 (0.015)	-0.018 (0.014)	42 383
Observations	84 834	42 441	42 393	

Notes: Fixed effects regressions for all pupils in all subjects combined and separately based on subject and gender, with standard errors clustered on class in 5th grade and school in 9th grade in parentheses. Significance levels are indicated as follows: + significant at 10 per cent; * significant at 5 per cent; ** significant at 1 per cent. Dependent variable is standardised test scores in national tests in 9th grade, and the coefficient of interest is the coefficient on the dummy for having low-achieving peer(s) in 9th grade. Coefficients are from separate regressions, based on model four (4) in Table 3, where I include fixed effects for the interaction between class in 5th grade and school in 9th grade, and controls for individual and parental characteristics. The first column represents all pupils combined, the second represents boys and the third column represents girls. In Panel A, I look at both subjects combined, and in Panels B and C I look only at test scores in reading and maths, respectively. The "Observations" column and row include the number of observations within the different categories.

Table E.5
Robustness test using a threshold of 15 per cent - subject and parental education

	Mother's education			Father's education		
	Not completed upper secondary school	Completed upper secondary school	University Degree	Not completed upper secondary school	Completed upper secondary school	University Degree
<i>Panel A. Both subjects</i>						
Having "low-achieving" peer(s)	-0.017 (0.019)	0.016 (0.014)	0.001 (0.011)	-0.010 (0.018)	0.001 (0.012)	0.005 (0.012)
<i>Panel C. Reading</i>						
Having "low-achieving" peer(s)	0.007 (0.029)	0.026 (0.020)	-0.002 (0.016)	-0.007 (0.028)	0.005 (0.018)	0.018 (0.017)
<i>Panel D. Maths</i>						
Having "low-achieving" peer(s)	-0.042 (0.025)	0.006 (0.018)	0.002 (0.015)	-0.011 (0.024)	-0.004 (0.016)	-0.009 (0.016)

Notes: Fixed effects regressions for subsamples of pupils in all subjects combined and separately based on subject and parental education, with standard errors clustered on class in 5th grade and school in 9th grade in parentheses. Significance levels are indicated as follows: + significant at 10 per cent; * significant at 5 per cent; ** significant at 1 per cent. Dependent variable is standardised test scores in national tests in 9th grade, and the coefficient of interest is the coefficient on the dummy for having low-achieving peer(s) in 9th grade. Coefficients are from separate regressions, based on model four (4) in Table 3, where I include fixed effects for the interaction between class in 5th grade and school in 9th grade, and controls for individual and parental characteristics. Parental education is specified in columns 1 to 6, with mother's education in columns 1 to 3 and father's education in columns 4 to 6. In Panel A, I look at both subjects combined, and in Panels B and C I look only at test scores in reading and maths, respectively.

Table E.6
Robustness test using a threshold of 15 per cent - subject and pupils' own position in the achievement distribution

	Pupils' own position in the achievement distribution				
	Below 20th percentile	20th to 40th percentile	40th to 60th percentile	60th to 80th percentile	Above 80th percentile
<i>Panel A. Both subjects</i>					
Having "low-achieving" peer(s)	-0.022 (0.038)	-0.039* (0.021)	0.039* (0.020)	0.006 (0.017)	0.011 (0.013)
<i>Panel C. Reading</i>					
Having "low-achieving" peer(s)	0.007 (0.059)	-0.035 (0.030)	0.052* (0.031)	0.022 (0.024)	0.018 (0.019)
<i>Panel D. Maths</i>					
Having "low-achieving" peer(s)	-0.046 (0.047)	-0.042 (0.029)	0.027 (0.026)	-0.013 (0.023)	0.006 (0.019)

Notes: Fixed effects regressions for subsamples of pupils in both subjects combined and separately based on subject and pupils' own position in the achievement distribution, with standard errors clustered on class in 5th grade and school in 9th grade in parentheses. Significance levels are indicated as follows: + significant at 10 per cent; * significant at 5 per cent; ** significant at 1 per cent. Dependent variable is standardised test scores in national tests in 9th grade, and the coefficient of interest is the coefficient on the dummy for having low-achieving peer(s) in 9th grade. Coefficients are from separate regressions, based on model four (4) in Table 3, where I include fixed effects for the interaction between class in 5th grade and school in 9th grade, and controls for individual and parental characteristics. Pupils' own position in the achievement distribution is specified in columns 1 to 5, with the lowest achieving pupils in column 1 and the highest achieving pupils in column 5. In Panel A, I look at both subjects combined, and in Panels B and C I look only at test scores in reading and maths, respectively.

Table E.7
Robustness test using only pupils who participated in the national tests - subject and gender

	All pupils	Gender		Observations
		Boys	Girls	
<i>Panel A. Both subjects</i>				
Having “low-achieving” peer(s)	-0.011 (0.008)	-0.010 (0.011)	-0.007 (0.011)	86 202
<i>Panel C. Reading</i>				
Having “low-achieving” peer(s)	-0.015 (0.012)	-0.017 (0.018)	-0.011 (0.016)	42 805
<i>Panel D. Maths</i>				
Having “low-achieving” peer(s)	-0.006 (0.010)	-0.004 (0.015)	-0.004 (0.014)	43 397
Observations	86 202	43 044	43 158	

Notes: Fixed effects regressions for pupils who participated in the national tests in both subjects combined and separately based on subject and gender, with standard errors clustered on class in 5th grade and school in 9th grade in parentheses. Significance levels are indicated as follows: + significant at 10 per cent; * significant at 5 per cent; ** significant at 1 per cent. Dependent variable is standardised test scores in national tests in 9th grade, and the coefficient of interest is the coefficient on the dummy for having low-achieving peer(s) in 9th grade. Coefficients are from separate regressions, based on model four (4) in Table 3, where I include fixed effects for the interaction between class in 5th grade and school in 9th grade, and controls for individual and parental characteristics. The first column represents all pupils combined, the second represents boys and the third column represents girls. In Panel A, I look at both subjects combined, and in Panels B and C I look only at test scores in reading and maths, respectively. The “Observations” column and row include the number of observations within the different categories.

Table E.8
Robustness test using only pupils who participated in the national tests - subject and parental education

	Mother’s education			Father’s education		
	Not completed upper secondary school	Completed upper secondary school	University Degree	Not completed upper secondary school	Completed upper secondary school	University Degree
<i>Panel A. Both subjects</i>						
Having “low-achieving” peer(s)	-0.001 (0.018)	-0.013 (0.014)	-0.005 (0.012)	-0.008 (0.017)	-0.017 (0.012)	-0.018 (0.013)
<i>Panel C. Reading</i>						
Having “low-achieving” peer(s)	-0.000 (0.026)	-0.016 (0.020)	-0.018 (0.017)	-0.032 (0.026)	-0.020 (0.018)	-0.006 (0.020)
<i>Panel D. Maths</i>						
Having “low-achieving” peer(s)	-0.004 (0.023)	-0.010 (0.018)	0.010 (0.016)	0.019 (0.022)	-0.014 (0.016)	-0.032* (0.017)

Notes: Fixed effects regressions for pupils who participated in the national tests in both subjects combined and separately based on subject and parental education, with standard errors clustered on class in 5th grade and school in 9th grade in parentheses. Significance levels are indicated as follows: + significant at 10 per cent; * significant at 5 per cent; ** significant at 1 per cent. Dependent variable is standardised test scores in national tests in 9th grade, and the coefficient of interest is the coefficient on the dummy for having low-achieving peer(s) in 9th grade. Coefficients are from separate regressions, based on model four (4) in Table 3, where I include fixed effects for the interaction between class in 5th grade and school in 9th grade, and controls for individual and parental characteristics. Parental education is specified in columns 1 to 6, with mother’s education in columns 1 to 3 and father’s education in columns 4 to 6. In Panel A, I look at both subjects combined, and in Panels B and C I look only at test scores in reading and maths, respectively.

Table E.9
Robustness test using only pupils who participated in the national tests - subject and pupils' own position in the achievement distribution

	Pupils' own position in the achievement distribution				
	Below 20th percentile	20th to 40th percentile	40th to 60th percentile	60th to 80th percentile	Above 80th percentile
<i>Panel A. Both subjects</i>					
Having "low-achieving" peer(s)	-0.052* (0.029)	-0.019 (0.021)	0.027 (0.020)	-0.024 (0.018)	-0.002 (0.015)
<i>Panel C. Reading</i>					
Having "low-achieving" peer(s)	-0.005 (0.045)	-0.062* (0.031)	0.028 (0.030)	-0.001 (0.026)	0.011 (0.022)
<i>Panel D. Maths</i>					
Having "low-achieving" peer(s)	-0.093* (0.038)	0.030 (0.029)	0.023 (0.026)	-0.043* (0.024)	-0.015 (0.020)

Notes: Fixed effects regressions for pupils who participated in the national tests in all subjects combined and separately based on subject and pupils' own position in the achievement distribution, with standard errors clustered on class in 5th grade and school in 9th grade in parentheses. Significance levels are indicated as follows: + significant at 10 per cent; * significant at 5 per cent; ** significant at 1 per cent. Dependent variable is standardised test scores in national tests in 9th grade, and the coefficient of interest is the coefficient on the dummy for having low-achieving peer(s) in 9th grade. Coefficients are from separate regressions, based on model four (4) in Table 3, where I include fixed effects for the interaction between class in 5th grade and school in 9th grade, and controls for individual and parental characteristics. Pupils' own position in the achievement distribution is specified in columns 1 to 5, with the lowest achieving pupils in column 1 and the highest achieving pupils in column 5. In Panel A, I look at both subjects combined, and in Panels B and C I look only at test scores in reading and maths, respectively.

Table E.10
Robustness test using pupils who are exempt from participating in the national tests as low-achievers - subject and gender

	All pupils	Gender		Observations
		Boys	Girls	
<i>Panel A. Both subjects</i>				
Having "low-achieving" peer(s)	-0.010 (0.007)	-0.005 (0.010)	-0.017* (0.010)	87 613
<i>Panel C. Reading</i>				
Having "low-achieving" peer(s)	-0.006 (0.010)	0.002 (0.016)	-0.016 (0.014)	43 681
<i>Panel D. Maths</i>				
Having "low-achieving" peer(s)	-0.015 (0.009)	-0.013 (0.014)	-0.018 (0.013)	43 613
Observations	87 613	43 810	43 803	

Notes: Fixed effects regressions for all pupils combined and separate based on subject and gender, with standard errors clustered on class in 5th grade and school in 9th grade in parentheses. Significance levels are indicated as follows: + significant at 10 per cent; * significant at 5 per cent; ** significant at 1 per cent. Dependent variable is standardised test scores in national tests in 9th grade, and the coefficient of interest is the coefficient on the dummy for having low-achieving peer(s) in 9th grade. Low-achievers are defined as pupils who are exempt from participating in the national tests in at least one subject. Coefficients are from separate regressions, based on model four (4) in Table 3, where I include fixed effects for the interaction between class in 5th grade and school in 9th grade, and controls for individual and parental characteristics. The first column represents all pupils combined, the second represents boys and the third column represents girls. In Panel A, I look at both subjects combined, and in Panels B and C I look only at test scores in reading and maths, respectively. The "Observations" column and row include the number of observations within the different categories.

Table E.11
Robustness test using pupils who are exempt from participating in the national tests as low-achievers - subject and parental education

	Mother's education			Father's education		
	Not completed upper secondary school	Completed upper secondary school	University Degree	Not completed upper secondary school	Completed upper secondary school	University Degree
<i>Panel A. Both subjects</i>						
Having "low-achieving" peer(s)	-0.045** (0.017)	0.012 (0.013)	-0.014 (0.011)	-0.018 (0.016)	-0.017 (0.011)	-0.004 (0.012)
<i>Panel C. Reading</i>						
Having "low-achieving" peer(s)	-0.046* (0.025)	0.011 (0.018)	-0.014 (0.015)	-0.008 (0.024)	-0.021 (0.016)	0.001 (0.017)
<i>Panel D. Maths</i>						
Having "low-achieving" peer(s)	-0.044* (0.022)	0.010 (0.017)	-0.013 (0.015)	-0.027 (0.021)	-0.016 (0.015)	-0.010 (0.017)

Notes: Fixed effects regressions for all pupils combined and separate based on subject and parental education, with standard errors clustered on class in 5th grade and school in 9th grade in parentheses. Significance levels are indicated as follows: + significant at 10 per cent; * significant at 5 per cent; ** significant at 1 per cent. Dependent variable is standardised test scores in national tests in 9th grade, and the coefficient of interest is the coefficient on the dummy for having low-achieving peer(s) in 9th grade. Low-achievers are defined as pupils who are exempt from participating in the national tests in at least one subject. Coefficients are from separate regressions, based on model four (4) in Table 3, where I include fixed effects for the interaction between class in 5th grade and school in 9th grade, and controls for individual and parental characteristics. Parental education is specified in columns 1 to 3, with mother's education in columns 1 to 3 and father's education in columns 4 to 6. In Panel A, I look at both subjects combined, and in Panels B and C I look only at test scores in reading and maths, respectively.

Table E.12
Robustness test using pupils who are exempt from participating in the national tests as low-achievers - subject and pupils' own position in the achievement distribution

	Pupils' own position in the achievement distribution				
	Below 20th percentile	20th to 40th percentile	40th to 60th percentile	60th to 80th percentile	Above 80th percentile
<i>Panel A. Both subjects</i>					
Having "low-achieving" peer(s)	-0.013 (0.026)	-0.069** (0.019)	-0.010 (0.019)	0.009 (0.016)	0.017 (0.013)
<i>Panel C. Reading</i>					
Having "low-achieving" peer(s)	-0.015 (0.040)	-0.059* (0.027)	0.000 (0.027)	0.013 (0.024)	0.017 (0.018)
<i>Panel D. Maths</i>					
Having "low-achieving" peer(s)	-0.012 (0.032)	-0.080** (0.026)	-0.018 (0.025)	0.003 (0.022)	0.016 (0.020)

Notes: Fixed effects regressions for all pupils combined and separate based on subject and pupils' own position in the achievement distribution, with standard errors clustered on class in 5th grade and school in 9th grade in parentheses. Significance levels are indicated as follows: + significant at 10 per cent; * significant at 5 per cent; ** significant at 1 per cent. Dependent variable is standardised test scores in national tests in 9th grade, and the coefficient of interest is the coefficient on the dummy for having low-achieving peer(s) in 9th grade. Low-achievers are defined as pupils who are exempt from participating in the national tests in at least one subject. Coefficients are from separate regressions, based on model four (4) in Table 3, where I include fixed effects for the interaction between class in 5th grade and school in 9th grade, and controls for individual and parental characteristics. Pupils' own position in the achievement distribution is specified in columns 1 to 5, with the lowest achieving pupils in column 1 and the highest achieving pupils in column 5. In Panel A, I look at both subjects combined, and in Panels B and C I look only at test scores in reading and maths, respectively.

Table E.13
Robustness test focusing on schools with few low-achievers - subject and gender

	Gender			Observations
	All pupils	Boys	Girls	
<i>Panel A. Both subjects</i>				
Having "low-achieving" peer(s)	-0.012 (0.009)	-0.004 (0.013)	-0.016 (0.012)	55 903
<i>Panel C. Reading</i>				
Having "low-achieving" peer(s)	-0.011 (0.012)	0.007 (0.019)	-0.024 (0.017)	27 696
<i>Panel D. Maths</i>				
Having "low-achieving" peer(s)	-0.014 (0.012)	-0.013 (0.017)	-0.010 (0.017)	28 207
Observations	55 903	27 938	27 965	

Notes: Fixed effects regressions for all pupils in both subjects combined and separately based on subject and gender, with standard errors clustered on class in 5th grade and school in 9th grade in parentheses. Significance levels are indicated as follows: + significant at 10 per cent; * significant at 5 per cent; ** significant at 1 per cent. Dependent variable is standardised test scores in national tests in 9th grade, and the coefficient of interest is the coefficient on the dummy for having low-achieving peer(s) in 9th grade. Coefficients are from separate regressions, based on model four (4) in Table 3, where I include fixed effects for the interaction between class in 5th grade and school in 9th grade, and controls for individual and parental characteristics. The first column represents all pupils combined, the second represents boys and the third column represents girls. In Panel A, I look at both subjects combined, and in Panels B and C I look only at test scores in reading and maths, respectively. The "Observations" column and row include the number of observations within the different categories.

Table E.14
Robustness test focusing on schools with few low-achievers - subject and parental education

	Mother's education			Father's education		
	Not completed upper secondary school	Completed upper secondary school	University Degree	Not completed upper secondary school	Completed upper secondary school	University Degree
<i>Panel A. Both subjects</i>						
Having "low-achieving" peer(s)	-0.014 (0.021)	-0.016 (0.016)	-0.028* (0.013)	-0.001 (0.020)	-0.029* (0.013)	-0.017 (0.014)
<i>Panel C. Reading</i>						
Having "low-achieving" peer(s)	-0.015 (0.031)	-0.023 (0.023)	-0.034* (0.018)	-0.012 (0.031)	-0.033* (0.020)	-0.003 (0.022)
<i>Panel D. Maths</i>						
Having "low-achieving" peer(s)	-0.015 (0.028)	-0.009 (0.021)	-0.023 (0.018)	0.015 (0.026)	-0.026 (0.018)	-0.032* (0.019)

Notes: Fixed effects regressions for subsamples of pupils in both subjects combined and separately based on subject and parental education, with standard errors clustered on class in 5th grade and school in 9th grade in parentheses. Significance levels are indicated as follows: + significant at 10 per cent; * significant at 5 per cent; ** significant at 1 per cent. Dependent variable is standardised test scores in national tests in 9th grade, and the coefficient of interest is the coefficient on the dummy for having low-achieving peer(s) in 9th grade. Coefficients are from separate regressions, based on model four (4) in Table 3, where I include fixed effects for the interaction between class in 5th grade and school in 9th grade, and controls for individual and parental characteristics. Parental education is specified in columns 1 to 6, with mother's education in columns 1 to 3 and father's education in columns 4 to 6. In Panel A, I look at both subjects combined, and in Panels B and C I look only at test scores in reading and maths, respectively.

Table E.15

Robustness test focusing on schools with few low-achievers - subject and pupils' own position in the achievement distribution

	Pupils' own position in the achievement distribution				
	Below 20th percentile	20th to 40th percentile	40th to 60th percentile	60th to 80th percentile	Above 80th percentile
<i>Panel A. Both subjects</i>					
Having "low-achieving" peer(s)	-0.027 (0.035)	-0.038 (0.024)	0.006 (0.022)	0.001 (0.019)	-0.009 (0.016)
<i>Panel C. Reading</i>					
Having "low-achieving" peer(s)	0.014 (0.053)	-0.066* (0.033)	0.021 (0.034)	0.007 (0.027)	-0.004 (0.023)
<i>Panel D. Maths</i>					
Having "low-achieving" peer(s)	-0.070 (0.045)	-0.009 (0.033)	-0.006 (0.029)	-0.004 (0.026)	-0.014 (0.023)

Notes: Fixed effects regressions for subsamples of pupils in both subjects combined and separately based on subject and pupils' own position in the achievement distribution, with standard errors clustered on class in 5th grade and school in 9th grade in parentheses. Significance levels are indicated as follows: + significant at 10 per cent; * significant at 5 per cent; ** significant at 1 per cent. Dependent variable is standardised test scores in national tests in 9th grade, and the coefficient of interest is the coefficient on the dummy for having low-achieving peer(s) in 9th grade. Coefficients are from separate regressions, based on model four (4) in Table 3, where I include fixed effects for the interaction between class in 5th grade and school in 9th grade, and controls for individual and parental characteristics. Pupils' own position in the achievement distribution is specified in columns 1 to 5, with the lowest achieving pupils in column 1 and the highest achieving pupils in column 5. In Panel A, I look at both subjects combined, and in Panels B and C I look only at test scores in reading and maths, respectively.

Table E.16

Robustness test using 8th grade test scores - subject and gender

	All pupils	Gender		Observations
		Boys	Girls	
<i>Panel A. All subjects</i>				
Having "low-achieving" peer(s)	-0.008* (0.004)	-0.006 (0.006)	-0.010+ (0.006)	219 613
<i>Panel B. English</i>				
Having "low-achieving" peer(s)	-0.006 (0.010)	-0.004 (0.014)	-0.008 (0.014)	43 563
<i>Panel C. Reading</i>				
Having "low-achieving" peer(s)	-0.002 (0.007)	-0.004 (0.011)	0.002 (0.010)	88 081
<i>Panel D. Maths</i>				
Having "low-achieving" peer(s)	-0.016* (0.006)	-0.008 (0.009)	-0.024** (0.009)	87 969
Observations	219 613	109 514	110 099	

Notes: Fixed effects regressions for all pupils in all subjects combined and separately based on subject and gender, with standard errors clustered on class in 5th grade and school in 8th grade in parentheses. Significance levels are indicated as follows: + significant at 10 per cent; * significant at 5 per cent; ** significant at 1 per cent. Dependent variable is standardised test scores in national tests in 8th grade, and the coefficient of interest is the coefficient on the dummy for having low-achieving peer(s) in 8th grade. Coefficients are from separate regressions, based on model four (4) in Table 3, where I include fixed effects for the interaction between class in 5th grade and school in 8th grade, and controls for individual and parental characteristics. The first column represents all pupils combined, the second represents boys and the third column represents girls. In Panel A, I look at all subjects combined, and in Panels B, C and D I look only at test scores in English, reading and maths, respectively. The "Observations" column and row include the number of observations within the different categories.

Table E.17
Robustness test using 8th grade test scores - subject and parental education

	Mother's education			Father's education		
	Not completed upper secondary school	Completed upper secondary school	University Degree	Not completed upper secondary school	Completed upper secondary school	University Degree
<i>Panel A. All subjects</i>						
Having "low-achieving" peer(s)	-0.007 (0.010)	-0.011 (0.008)	0.001 (0.006)	-0.005 (0.010)	-0.005 (0.007)	-0.005 (0.007)
<i>Panel B. English</i>						
Having "low-achieving" peer(s)	0.005 (0.025)	-0.019 (0.017)	0.010 (0.015)	-0.008 (0.023)	-0.005 (0.016)	-0.005 (0.017)
<i>Panel C. Reading</i>						
Having "low-achieving" peer(s)	-0.017 (0.017)	0.003 (0.013)	0.007 (0.010)	-0.013 (0.016)	0.001 (0.011)	0.011 (0.012)
<i>Panel D. Maths</i>						
Having "low-achieving" peer(s)	-0.000 (0.015)	-0.021 ⁺ (0.011)	-0.008 (0.010)	0.005 (0.014)	-0.010 (0.010)	-0.023 ⁺ (0.011)

Notes: Fixed effects regressions for subsamples of pupils in all subjects combined and separately based on subject and parental education, with standard errors clustered on class in 5th grade and school in 8th grade in parentheses. Significance levels are indicated as follows: + significant at 10 per cent; * significant at 5 per cent; ** significant at 1 per cent. Dependent variable is standardised test scores in national tests in 8th grade, and the coefficient of interest is the coefficient on the dummy for having low-achieving peer(s) in 8th grade. Coefficients are from separate regressions, based on model four (4) in Table 3, where I include fixed effects for the interaction between class in 5th grade and school in 8th grade, and controls for individual and parental characteristics. Parental education is specified in columns 1 to 6, with mother's education in columns 1 to 3 and father's education in columns 4 to 6. In Panel A, I look at all subjects combined, and in Panels B, C and D I look only at test scores in English, reading and maths, respectively.

Table E.18
Robustness test using 8th grade test scores - subject and pupils' own position in the achievement distribution

	Pupils' own position in the achievement distribution				
	Below 20th percentile	20th to 40th percentile	40th to 60th percentile	60th to 80th percentile	Above 80th percentile
<i>Panel A. All subjects</i>					
Having "low-achieving" peer(s)	-0.044 ⁺ (0.017)	-0.000 (0.012)	-0.006 (0.011)	-0.014 (0.010)	-0.016 ⁺ (0.009)
<i>Panel B. English</i>					
Having "low-achieving" peer(s)	-0.028 (0.038)	-0.035 (0.027)	0.006 (0.026)	-0.004 (0.025)	-0.029 (0.019)
<i>Panel C. Reading</i>					
Having "low-achieving" peer(s)	-0.036 (0.029)	0.021 (0.019)	-0.006 (0.018)	-0.004 (0.016)	0.001 (0.014)
<i>Panel D. Maths</i>					
Having "low-achieving" peer(s)	-0.056 ⁺ (0.022)	-0.008 (0.017)	-0.011 (0.016)	-0.029 ⁺ (0.016)	-0.026 ⁺ (0.014)

Notes: Fixed effects regressions for subsamples of pupils in all subjects combined and separately based on subject and pupils' own position in the achievement distribution, with standard errors clustered on class in 5th grade and school in 8th grade in parentheses. Significance levels are indicated as follows: + significant at 10 per cent; * significant at 5 per cent; ** significant at 1 per cent. Dependent variable is standardised test scores in national tests in 8th grade, and the coefficient of interest is the coefficient on the dummy for having low-achieving peer(s) in 8th grade. Coefficients are from separate regressions, based on model four (4) in Table 3, where I include fixed effects for the interaction between class in 5th grade and school in 8th grade, and controls for individual and parental characteristics. Pupils' own position in the achievement distribution is specified in columns 1 to 5, with the lowest achieving pupils in column 1 and the highest achieving pupils in column 5. In Panel A, I look at all subjects combined, and in Panels B, C and D I look only at test scores in English, reading and maths, respectively.

Table E.19
Robustness test using only non-immigrant pupils - subject and gender

	All pupils	Gender		Observations
		Boys	Girls	
<i>Panel A. Both subjects</i>				
Having “low-achieving” peer(s)	-0.014* (0.008)	-0.008 (0.013)	-0.017 (0.012)	62 840
<i>Panel C. Reading</i>				
Having “low-achieving” peer(s)	-0.016 (0.012)	-0.015 (0.019)	-0.013 (0.017)	31 404
<i>Panel D. Maths</i>				
Having “low-achieving” peer(s)	-0.012 (0.011)	-0.000 (0.017)	-0.022 (0.016)	31 436
Observations	62 840	31 587	31 253	

Notes: Fixed effects regressions for non-immigrant pupils in both subjects combined and separately based on subject and gender, with standard errors clustered on class in 5th grade and school in 9th grade in parentheses. Significance levels are indicated as follows: + significant at 10 per cent; * significant at 5 per cent; ** significant at 1 per cent. Dependent variable is standardised test scores in national tests in 9th grade, and the coefficient of interest is the coefficient on the dummy for having non-immigrant low-achieving peer(s) in 9th grade. Coefficients are from separate regressions, based on model four (4) in Table 3, where I include fixed effects for the interaction between class in 5th grade and school in 9th grade, and controls for individual and parental characteristics. The first column represents all non-immigrant pupils combined, the second represents non-immigrant boys and the third column represents non-immigrant girls. In Panel A, I look at both subjects combined, and in Panels B and C I look only at test scores in reading and maths, respectively. The “Observations” column and row include the number of observations within the different categories.

Table E.20
Robustness test using only non-immigrant pupils - subject and parental education

	Mother’s education			Father’s education		
	Not completed upper secondary school	Completed upper secondary school	University Degree	Not completed upper secondary school	Completed upper secondary school	University Degree
<i>Panel A. Both subjects</i>						
Having “low-achieving” peer(s)	0.009 (0.021)	-0.034* (0.015)	-0.023* (0.013)	-0.022 (0.021)	-0.016 (0.013)	-0.013 (0.015)
<i>Panel C. Reading</i>						
Having “low-achieving” peer(s)	0.005 (0.031)	-0.046* (0.021)	-0.021 (0.019)	-0.005 (0.032)	-0.022 (0.019)	-0.017 (0.022)
<i>Panel D. Maths</i>						
Having “low-achieving” peer(s)	0.011 (0.028)	-0.025 (0.020)	-0.024 (0.018)	-0.042 (0.027)	-0.011 (0.017)	-0.011 (0.022)

Notes: Fixed effects regressions for subsamples of non-immigrant pupils in both subjects combined and separately based on subject and parental education, with standard errors clustered on class in 5th grade and school in 9th grade in parentheses. Significance levels are indicated as follows: + significant at 10 per cent; * significant at 5 per cent; ** significant at 1 per cent. Dependent variable is standardised test scores in national tests in 9th grade, and the coefficient of interest is the coefficient on the dummy for having non-immigrant low-achieving peer(s) in 9th grade. Coefficients are from separate regressions, based on model four (4) in Table 3, where I include fixed effects for the interaction between class in 5th grade and school in 9th grade, and controls for individual and parental characteristics. Parental education is specified in columns 1 to 6, with mother’s education in columns 1 to 3 and father’s education in columns 4 to 6. In Panel A, I look at both subjects combined, and in Panels B and C I look only at test scores in reading and maths, respectively.

Table E.21
Robustness test using only non-immigrant pupils - subject and pupils' own position in the achievement distribution

	Pupils' own position in the achievement distribution				
	Below 20th percentile	20th to 40th percentile	40th to 60th percentile	60th to 80th percentile	Above 80th percentile
<i>Panel A. Both subjects</i>					
Having "low-achieving" peer(s)	-0.040 (0.035)	-0.003 (0.024)	-0.004 (0.022)	-0.024 (0.019)	0.008 (0.017)
<i>Panel C. Reading</i>					
Having "low-achieving" peer(s)	-0.026 (0.056)	-0.012 (0.035)	0.001 (0.032)	-0.021 (0.027)	-0.006 (0.023)
<i>Panel D. Maths</i>					
Having "low-achieving" peer(s)	-0.051 (0.042)	0.003 (0.034)	-0.006 (0.031)	-0.026 (0.027)	0.020 (0.025)

Notes: Fixed effects regressions for subsamples of non-immigrant pupils in all subjects combined and separately based on subject and pupils' own position in the achievement distribution, with standard errors clustered on class in 5th grade and school in 9th grade in parentheses. Significance levels are indicated as follows: + significant at 10 per cent; * significant at 5 per cent; ** significant at 1 per cent. Dependent variable is standardised test scores in national tests in 9th grade, and the coefficient of interest is the coefficient on the dummy for having non-immigrant low-achieving peer(s) in 9th grade. Coefficients are from separate regressions, based on model four (4) in Table 3, where I include fixed effects for the interaction between class in 5th grade and school in 9th grade, and controls for individual and parental characteristics. Pupils' own position in the achievement distribution is specified in columns 1 to 5, with the lowest achieving pupils in column 1 and the highest achieving pupils in column 5. In Panel A, I look at both subjects combined, and in Panels B and C I look only at test scores in reading and maths, respectively.

References

- Ainscow, M., César, M., 2006. Inclusive education ten years after Salamanca: Setting the agenda. *Eur. J. Psych. Educ.* 21 (3), 231–238.
- Aizer, A., 2008. Peer effects and human capital accumulation: The externalities of ADD. NBER Working Paper No. 14354. National Bureau of Economic Research, Cambridge, MA. <https://www.doi.org/10.3386/w14354>.
- Angrist, J.D., 2014. The perils of peer effects. *Labour Econ.* 30, 98–108.
- Bedard, K., Dhuey, E., 2006. The persistence of early childhood maturity: international evidence of long-run age effects. *Q. J. Econ.* 121 (4), 1437–1472.
- Boarini, R., 2009. Making the most of Norwegian Schools. OECD Economics Department Working Papers, No. 661. OECD Publishing, Paris. <https://doi.org/10.1787/227333554837>.
- Burke, M.A., Sass, T.R., 2013. Classroom peer effects and student achievement. *J. Labor Econ.* 31 (1), 51–82.
- Carrell, S.E., Fullerton, R.L., West, J.E., 2009. Does your cohort matter? Measuring peer effects in college achievement. *J. Labor Econ.* 27 (3), 439–464.
- Carrell, S.E., Hoekstra, M., Kuka, E., 2016. The long-run effects of disruptive peers. NBER Working Paper No. 22042. National Bureau of Economic Research, Cambridge, MA. <https://www.doi.org/10.3386/w22042>.
- Carrell, S.E., Hoekstra, M.L., 2010. Externalities in the classroom: how children exposed to domestic violence affect everyone's kids. *Am. Econ. J. Appl. Econ.* 2 (1), 211–228.
- European Agency for Development in Special Needs Education, 2013. Organisation of provision to support inclusive education Literature review. Odense, Denmark. https://www.european-agency.org/sites/default/files/organisation-of-provision-to-support-inclusive-education-2013-literature-review_Organisation-of-Provision-Literature-Review.pdf
- Eurydice (2005). Key data on education in Europe 2005. European Commission, Eurydice, Eurostat. Luxembourg: Office for Official Publications of the European Communities. http://www.indire.it/lucabas/lkmw_file/eurydice/Key_Data_2005_EN.pdf.
- Feld, J., Zölitz, U., 2017. Understanding peer effects: on the nature, estimation, and channels of peer effects. *J. Labor Econ.* 35 (2), 387–428.
- Figlio, D.N., 2007. Boys named sue: disruptive children and their peers. *Educ. Financ. Policy* 2 (4), 376–394.
- Fruehwirth, J.C., 2014. Can achievement peer effect estimates inform policy? A view from inside the black box. *Rev. Econ. Stati.* 96 (3), 514–523.
- Gibbons, S. and Telhaj, S., 2008. Peers and achievement in England's secondary schools. SERC Discussion Paper 1. London, UK. <http://eprints.lse.ac.uk/id/eprint/30854>.
- Hanushek, E.A., Kain, J.F., Markman, J.M., Rivkin, S.G., 2003. Does peer ability affect student achievement? *J. Appl. Econ.* 18 (5), 527–544.
- Hanushek, E.A., Kain, J.F., Rivkin, S.G., 2002. Inferring program effects for special populations: does special education raise achievement for students with disabilities? *Rev. Econ. Stat.* 84 (4), 584–599.
- Hoxby, C., 2000. Peer effects in the classroom: Learning from gender and race variation. NBER Working Paper No. 7867. National Bureau of Economic Research, Cambridge, MA. <https://www.doi.org/10.3386/w7867>.
- Hoxby, C. M., and Weingarth, G. (2005). Taking race out of the equation: School reassignment and the structure of peer effects (No. 7867). Unpublished. <https://www.pausd.org/sites/default/files/pdf-faqs/attachments/TakingRaceOutOfTheEquation.pdf>.
- Huber, K.D., Rosenfeld, J.G., Fiorello, C.A., 2001. The differential impact of inclusion and inclusive practices on high, average, and low achieving general education students. *Psychol. Sch.* 38 (6), 497–504.
- Imberman, S.A., Kugler, A.D., Sacerdote, B.L., 2012. Katrina's children: evidence on the structure of peer effects from hurricane evacuees. *Am. Econ. Rev.* 102 (5), 2048–2082.
- Kalambouka, A., Farrell, P., Dyson, A., Kaplan, I., 2007. The impact of placing pupils with special educational needs in mainstream schools on the achievement of their peers. *Educ. Res.* 49 (4), 365–382.
- Kristoffersen, J.H.G., Krægpøth, M.V., Nielsen, H.S., Simonsen, M., 2015. Disruptive school peers and student outcomes. *Econ. Educ. Rev.* 45, 1–13.
- Lavy, V., Paserman, M.D., Schlosser, A., 2012. Inside the black box of ability peer effects: evidence from variation in the proportion of low achievers in the classroom. *Econ. J.* 122 (559), 208–237.
- Lavy, V., Schlosser, A., 2011. Mechanisms and impacts of gender peer effects at school. *Am. Econ. J. Appl. Econ.* 3 (2), 1–33.
- Lavy, V., Silva, O., Weinhardt, F., 2012. The good, the bad, and the average: evidence on ability peer effects in schools. *J. Labor Econ.* 30 (2), 367–414.
- Lazear, E.P., 2001. Educational production. *The Quarterly Journal of Economics* 116 (3), 777–803. doi:10.1162/00335530152466232.
- Lillejord, S., 2015. Den lange veien mot en inkluderende skole. *Utdanningsforskning*. <https://utdanningsforskning.no/artikler/den-lange-veien-mot-en-inkluderende-skole/>
- Manski, C.F., 1993. Identification of endogenous social effects: the reflection problem. *Rev. Econ. Stud.* 60 (3), 531–542.
- Ministry of Education and Research (2017). *Veiledning om organisering av elevene*. Kunnskapsdepartementet https://www.regjeringen.no/contentassets/f94154aa3d2b491ba1ac2f7f658cb019/veiledning-om-organisering-av-elveene_oppdater-april-2017.pdf.
- Ministry of Education and Science, M.S., United Nations Educational Scientific and Cultural Organization, P.F., 1994. The Salamanca statement and framework for action on special needs education. Technical Report. World conference on special needs education: Access and Quality.
- Norwegian Directorate for Education and Training, 2018. *Nasjonale prøver*. Utdanningsdirektoratet. <https://www.udir.no/eksamen-og-prover/prover/nasjonale-prover/om-nasjonale-prover/>.
- Norwegian Directorate for Education and Training, 2016. *Har eleven rett til særskilt språkopplæring?* Utdanningsdirektoratet. <https://www.udir.no/laring-og-trivsel/minoritetsspraklige/regleverk-som-gjelder-spesielt-for-minoritetsspraklige/sarskilt-sprakopplaring/Hva-betyr-retten-til-sarskilt-sprakopplaring/>.
- Norwegian Directorate for Education and Training, 2018a. *Administrere nasjonale prøver*. Utdanningsdirektoratet. <https://www.udir.no/eksamen-og-prover/prover/nasjonale-prover/administrere-nasjonale-prover2/>.
- Norwegian Directorate for Education and Training, 2018b. *Spesialundervisning - rettigheter*. Utdanningsdirektoratet. <https://www.udir.no/laring-og-trivsel/sarskille-behov/spesialundervisning/spesialundervisning-til-deg-som-er-forelder/>.
- OECD, 2012. Special educational needs (SEN). <http://www.oecd.org/els/family/50325299.pdf>.
- OECD, 2017. *Education at a glance 2017: OECD indicators*. OECD Publishing, Paris. <https://doi.org/10.1787/eag-2017-en>.

- Ogden, T., 2014. Special needs education in Norway - the past, present, and future of the field. In: Cook, T.J.L.e. Bryan G., Tankersley, M. (Eds.), *Special Education Past, Present, and Future: Perspectives from the Field*. In: *Advances in Learning and Behavioural Disabilities*, 27. Emerald Group Publishing Limited, pp. 213–238.
- Ruijs, N.M., 2017. The impact of special needs students on classmate performance. *Econ. Educ. Rev.* 58, 15–31.
- Ruijs, N.M., Peetsma, T.T., 2009. Effects of inclusion on students with and without special educational needs reviewed. *Educ. Res. Rev.* 4 (2), 67–79.
- Ruijs, N.M., Van der Veen, I., Peetsma, T.T., 2010. Inclusive education and students without special educational needs. *Educ. Res.* 52 (4), 351–390.
- Sacerdote, B., 2011. Peer Effects in Education: How Might They Work, How Big Are They and How Much Do We Know Thus Far?, 3. Elsevier B.V., United Kingdom, pp. 249–277.
- Stinebrickner, Todd R., Ralph Stinebrickner (2001). Peer Effects among Students from Disadvantaged Backgrounds. CIBC Centre for Human Capital and Productivity. CIBC Working Papers. London, ON: Department of Economics, University of Western Ontario. https://ir.lib.uwo.ca/cgi/viewcontent.cgi?article=1030&context=economicscibc&sei-redir=1&referer=https%253A%252F%252Fscholar.google.com%252Fscholar%253Fhl%253Dno%2526as_sdt%253D0%25252C5%2526q%253DPeer%252Beffects%252Bamong%252Bstudents%252Bfrom%252Bdisadvantaged%252Bbackgrounds%2526btnG%253D#search=%22Peer%20effects%20among%20students%20from%20disadvantaged%20backgrounds%22.
- Strøm, B. (2004). Student achievement and birthday effects. Unpublished manuscript, Norwegian University of Science and Technology. <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.168.5995&rep=rep1&type=pdf>.
- The Education Act (Opplæringslova), 1998. Lov om grunnskolen og den videregående opplæringa (LOV-1998-07-17-61). <https://lovdata.no/dokument/NL/lov/1998-07-17-61>.

Paper II



Girls in STEM: the role of teacher mindset

May Linn Auestad*

Eric Bettinger*[†]

Mari Rege*

Ingeborg F. Solli*

April, 2020[‡]

Abstract

Women outnumber men at most college campuses around the world, but remain underrepresented in STEM fields. One explanation is that socio-emotional learning such as students' mindsets differ between males and females. In this study, we investigate the effects of having a “high growth mindset” teacher on math outcomes during high school. We find a positive effect on students' mindset, attitudes and expectations in relation to math. In addition, we find that students are more likely to choose advanced math courses in high school, a critical gateway for continuing towards a STEM career. We do not, however, find any effect on achievement. We find that socio-emotional learning matters for both males and females but not differentially.

*University of Stavanger

[†]Stanford University and NBER

[‡]The authors acknowledge data from the *Usay: Shaping students learning expectations* project funded by the Norwegian Research Council (grant number: 260407) led by Professor Mari Rege. We are thankful for helpful comments from David S. Yeager, Kjell Arne Brekke and Maximiliaan Willem Pierre Thijssen, participants of the Psychology and Education Workshop, participants of the Quantitative Forum at the University of Stavanger, and participants of the Ph.D. Workshop in Education Economics. Corresponding author's email: maylinna@gmail.com

1 Introduction

While outnumbering men at most college campuses throughout the world, women remain underrepresented in the core subjects of science, technology, engineering and math, often referred to as the STEM fields. Reducing the gender imbalance within these fields might lead to new and improved innovations (Kahn & Ginther, 2017), and reduce the gender wage gap, as these jobs generally pay higher salaries than non-STEM-related jobs. Previous research has looked at how to recruit more women into these fields, with papers focused on role models (Bettinger & Long, 2005), the gender of instructors (Carrell et al., 2010), the characteristics of colleges (Griffith, 2010), financial aid (Sjoquist & Winters, 2015) and high school preparation (Card & Payne, 2017).

Interest gaps in STEM-related subjects start early in students' careers. Girls and boys show different levels of engagement in math as early as age six (Petersen & Hyde, 2017). The gender gap continues through middle school and manifests itself in high school as women take less rigorous mathematics courses (e.g. Xie & Shauman, 2005). These gaps lead to gender differences in the likelihood of studying math-intensive fields in college. In the United States for instance, women make up 69 percent of bachelor degrees in science-related fields that are not mathematically intensive¹ and only 27 percent of bachelor degrees in mathematically intensive fields² (Kahn & Ginther, 2017). Despite the difference in representation, several studies report only minimal differences in terms of average achievement in the subjects most important for a STEM carrier later on (Petersen & Hyde, 2014; Lindberg et al., 2010; Hyde et al., 2008).

Related literature in social psychology has focused primarily on students' attitudes and beliefs. The mindset literature pioneered by Dweck (1986, 2006) focuses on students' beliefs about their own ability to learn. Students with a "growth mindset" believe that intelligence is malleable and that strenuous effort can improve one's capacity and ability in a given area. By contrast, students with a "fixed mindset" believe that effort does not lead to subsequent improvement. In economic terms, fixed mindset students feel that they have an endowment of understanding that cannot be improved or altered.

Economists are just beginning to turn their attention to the growth mindset. Well-structured randomized experiments in social psychology and economics have repeatedly shown that educational interventions can cause students to move from a fixed to a growth mindset and that this shift results in changes in challenge-seeking behavior and improved academic performance (e.g., Bettinger et al., 2018; Alan & Ertac, 2017; Andersen & Nielsen, 2016).³ However, while the growth mindset can be taught through a focused curriculum (Bettinger et al., 2018; Yeager, Walton, et al., 2016; Paunesku et al., 2015; Aronson et al., 2002), many have postulated that teachers are a primary mechanism by

¹Life sciences, psychology and social sciences (excluding economics).

²Geo-sciences, engineering, economics, math/computer science, and physical science.

³Recent work by Dobronyi et al. (2019) has failed to replicate the findings from growth mindset interventions.

which a mindset is developed (e.g. Dweck, 2014). Teachers might in fact be the key mechanism by which growth mindset interventions may be scaled; however, to date, there have been few rigorous studies of the relationship between teacher and student mindset.

The causal impact of teachers, however, is an oft-explored theme in the academic literature (e.g. Hanushek, 1971; Rockoff, 2004; Chetty et al., 2014a,b). While researchers have been able to identify the value added that teachers provide in terms of academic achievement (e.g. Rockoff, 2004; Chetty et al., 2014a) and earnings (e.g. Chetty et al., 2014b), there is still significant discussion as to whether value added arises from the teacher’s impact on the cognitive or social-emotional attributes of students. Moreover, several studies have demonstrated the importance of children’s social-emotional development for academic achievement (Heckman et al., 2006; Kautz et al., 2014). Teachers’ value added may be the result of their impact on a variety of student characteristics (Araujo et al., 2016; Blazar & Kraft, 2017; Kraft, 2019), including mindset.

Prior work has suggested that mindset can be a contributing factor in subsequent course selection, especially among girls (Blackwell et al., 2007; Good et al., 2012; Nix et al., 2015; Perez-Felkner et al., 2017). In addition, some studies suggest that one reason for different participation among boys and girls might be that girls are not thriving within STEM fields due to stereotypes, culture, role models, competition, risk aversion or interests (Dweck, 1986; Kahn & Ginther, 2017). They also suggest that motivational interventions or role models, like teachers, might help reduce the gap that exists. Other studies present suggestive evidence that teacher and peer support are more important for girls, especially in math and more advanced courses (Riegle-Crumb et al., 2006; Beilock et al., 2010; Samuelsson & Samuelsson, 2016).

We attempt to link these important strands of scholarship. Our study examines attitudes towards math and participation in the most challenging math course in secondary schooling. In particular, we look at whether being exposed to a teacher with a relatively stronger growth mindset might explain students’ expectations and interest in math as well as their pursuit of the most rigorous course offerings in the subject. We do this by building on the existing literature on growth mindset and teacher value added and considering the extent to which students’ mindset, attitudes, challenge-seeking behavior and achievement were affected by their math teacher. In addition, because of suggestive evidence that the teacher might play a more important role for STEM participation among girls, we divide the sample based on gender and focus on girls in particular.

Our research takes advantage of a distinctive feature in the Norwegian educational system, specifically our identifying assumption that students are randomly assigned to classes, and thus teachers, within elementary and middle schools. In addition, all Norwegian students, regardless of past performance, generally have the same curriculum through middle school (until the end of 10th grade).⁴ The period of interest in the present study, specifically the transition to high school, is important, as this is the first instance of track-

⁴Except in electives.

ing in the Norwegian educational system, where students can choose between an academic or vocational track. Furthermore, first year of the academic track at high school, students can choose to pursue a rigorous or a less rigorous math track. This decision has consequential effects, since it impacts on the student's likelihood of studying STEM fields in college (Sells, 1980; Ma & Johnson, 2008).

We utilize a unique dataset with extensive survey data for a full student cohort from two of the largest counties in Norway. Survey data are matched with registry data on educational outcomes, class assignment, and family characteristics. The survey measures teacher mindset by using a student's peers' perceptions of their 10th grade math teacher. Notably, students' perceptions of their teacher's mindset may be correlated with several teacher characteristics (e.g. conscientiousness, optimism, ability to ask critical reasoning questions). Hence, while we attempt to align our definition of growth mindset with earlier literature, we acknowledge that our "teacher mindset" measure may also reflect other teacher characteristics, including being a more effective instructor.

Our results suggest that teachers with more of a growth mindset cause their students to develop a similar mindset. The effects are particularly strong when examining students' attitudes about math. We also examine students' academic choices and performance. We do not find evidence that the math teacher's mindset leads to increased enrollment in the academic track or performance in math. We do, however, find that students exposed to a math teacher with more of a growth mindset are more likely to engage in the most rigorous math curriculum in high school. Finally, we find no differential effects across gender: being exposed to a teacher with more of a growth mindset is equally beneficial for boys and girls.

The rest of the paper is organized as follows. First, in [Section 2](#) we present previous literature and our conceptual framework, and briefly explain the Norwegian educational system in [Section 3](#). In [Sections 4](#) and [5](#) we present the empirical strategy and data used to investigate the effects. The results follow in [Section 6](#), before finally, we conclude in [Section 7](#).

2 The impact of teacher mindset: conceptual framework

In this study, we investigate the effect of the teacher's mindset on the students' mindsets, attitudes, challenge-seeking behavior and achievements. We do this by comparing students' outcomes with their 10th grade math teacher's mindset. Thus, the mechanism we test in this paper is the transmission of a growth mindset from a teacher to his or her students, and we look at the effect on other outcomes as well. [Figure 1](#) presents an illustration of our conceptual framework. The mechanism of interest is related to several complementary research strands.

First, there is a long tradition of identifying the impact of teachers on students (Hanushek, 1971; Murnane, 1975; Rockoff, 2004; Chetty et al., 2014a,b; Kraft, 2019). The value added literature systematically identifies teachers whose students excel more

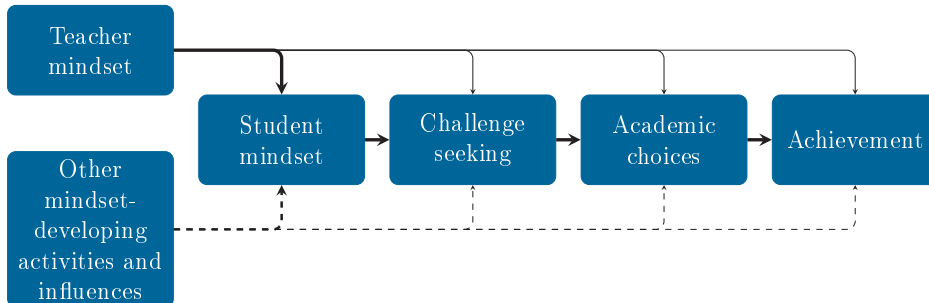


Figure 1: Conceptual framework

than expected, given the observable characteristics (parental education, gender, prior test scores, etc.). The literature, however, struggles to identify the specific mechanisms by which teachers affect students. Recent studies investigating teacher effects provide evidence that teachers affect not only performance, but also students' attitudes, behaviors and other social-emotional competencies (Blazar & Kraft, 2017; Kraft, 2019). A related strand of literature identifying best practices in teaching, focuses on a set of qualitative metrics on teaching style to identify mechanisms by which students learn from teachers (e.g. Grossman et al., 2013; Kane et al., 2011). However, identification is a challenge in many studies, since teachers are not usually randomly assigned to classes.

Another related strand of literature is the literature on mindset (or "implicit theory"). This literature defines two extreme types of mindset with a spectrum in between (e.g. Dweck & Leggett, 1988; Dweck, 2006; Yeager & Dweck, 2012). At one end, we find people with a growth mindset, who believe that talents and intelligence can be developed through effort, different strategies and perseverance. They are process-oriented and know that success requires hard work. People with a fixed mindset, at the other end of the spectrum, believe that talents and intelligence are predetermined traits - that some people are born smart, while others are not - and that there is nothing they can do about it.

There are several papers which demonstrate that exposing students to a growth mindset curriculum can increase the prevalence of the growth mindset, and have a positive effect on student outcomes, like achievement and challenge-seeking behavior (Mueller & Dweck, 1998; Good et al., 2003; Blackwell et al., 2007; Paunesku et al., 2015; Yeager, Romero, et al., 2016; Yeager, Walton, et al., 2016; Andersen & Nielsen, 2016; Bettinger et al., 2018; Yeager et al., 2019). However, there is not a consensus in the literature, as recent work by Dobronyi et al. (2019) has failed to replicate the findings from growth mindset interventions. While a growth mindset might be taught through a focused curriculum, it appears to be an inherent assumption in the literature that teachers can transmit mindset in distinct and effective ways (e.g. Dweck, 2006, 2014; Heggart, 2017). For example, Dweck (2006) outlines the importance of utilizing learning tasks that encourage a growth mindset. Heggart (2017) focuses on the importance of teachers' acquisition of a growth

mindset to encourage students to do likewise. However, there is limited evidence to date on the impact of teacher mindset on student outcomes (Brooks & Goldstein, 2008).

Furthermore, empirical evidence suggests that mindset might be a contributing factor in course selection, especially among girls (Blackwell et al., 2007; Good et al., 2012; Nix et al., 2015; Perez-Felkner et al., 2017). For example, girls with a more growth mindset might be more likely to choose more advanced courses than those with a more fixed mindset. Others suggest that attitudes towards math, in terms of interest and expectations, might be one of the reasons for different STEM participation between boys and girls (Bandura et al., 2001; Eccles, 1994; Petersen & Hyde, 2014). Moreover, some studies propose that one of the reasons for girls being less likely than boys to choose a STEM career might be that they are not thriving within STEM fields due to stereotypes, culture, role models, competition, risk aversion or interests (Dweck, 1986; Kahn & Ginther, 2017). Dweck (1986) and Kahn & Ginther (2017) also suggest that motivational interventions or role models, such as teachers, might help more girls to believe that they can succeed within math-intensive fields, and thereby reduce the gap that exists. Finally, some studies suggest that teacher and peer support are more important for girls, especially in math and more advanced courses (Riegle-Crumb et al., 2006; Beilock et al., 2010; Samuelsson & Samuelsson, 2016). Within our context, this could mean that, if exposed to a teacher with more of a growth mindset, girls in particular might be more likely to hold a growth mindset themselves. This in turn, could affect their response to challenges and setbacks, driving more girls towards a career within STEM.

3 Institutional context

3.1 Education in Norway

Elementary and middle schools in Norway provide an ideal place to examine the impact of mindset. Specifically, the uniform curriculum across schools and classrooms, and random assignment of students to teachers within schools, provide us with a unique opportunity to identify the impact of teacher mindset on student outcomes. Below we provide a brief description of the Norwegian educational system, the random assignment of students to teachers, and some specifics about the math teachers in the last year of middle school.

3.1.1 The Norwegian educational system

An illustration of the Norwegian educational system is provided in [Figure 2](#). Mandatory schooling in Norway is divided into primary school (elementary, 1st-7th grade) and lower secondary school (middle, 8th-10th grade). Children enroll at elementary school the calendar year they turn six (the Ministry of Education and Research, 1998). Non-compliance with this enrollment regulation is rare, which is also repeating or excelling a grade. Consequently, most students graduate from mandatory schooling the calendar year they turn sixteen. Public schools in Norway are free for all children to attend, and

only 4 percent attend private elementary or middle schools (Ministry of Education and Research, 1998; Statistics Norway, 2017a). At the end of middle school, students are randomly drawn to sit for one written and one oral externally graded exam. The written exam is either in Norwegian, English or math, which implies that one-third of students sit the math exam each year. The two exam grades and ten teacher-set grades constitute the grade point average from middle school (10th grade GPA).

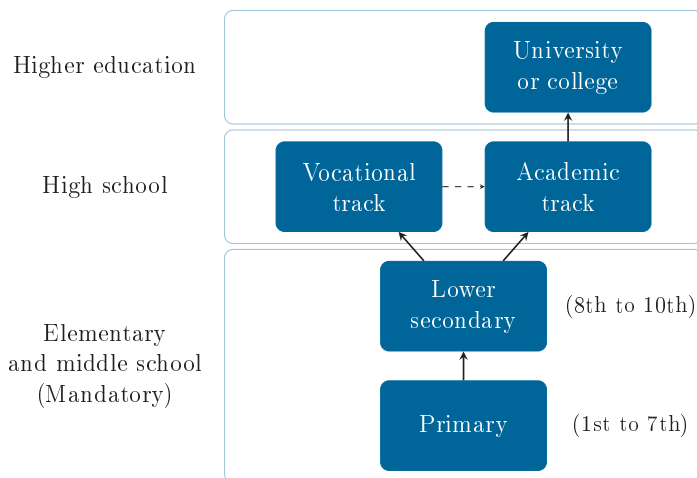


Figure 2: The Norwegian educational system

Elementary and middle schools are governed by the municipal administrations as district schools, and most children attend a school in their neighborhood. Middle schools tend to be larger than elementary schools, implying that several elementary schools feed into the same middle school.

After finishing middle school in 10th grade, all children have the right to proceed to upper secondary school (high school). When applying for high school, students can choose between several academic and vocational tracks. Vocational tracks consist of three to four years of education and training oriented around specific occupations, where the objective is to prepare students for the labor market. Academic tracks consist of three years of schooling and are intended to prepare students for university. Students graduating from the vocational tracks have the opportunity to attend university if they do an additional one-year of intensive academic training (Markussen & Gloppen, 2012). This means that students start making decisions in high school that will affect their possibilities later on. Another such decision is the choice between practical or theoretical math, hereafter referred to as regular and advanced math, if following the academic track. The choice to participate in the advanced math course in high school is considered by many to be a critical gateway to continuing towards a STEM career later on (Sells, 1980; Freeman & Aspray, 1999; Ma & Johnson, 2008). This also applies to the Norwegian context, as

there are many majors, primarily within the STEM fields, that require the student to have completed the advanced math course.

High schools in Norway are governed by the county administrations, and students can apply to attend any high school and educational program in their county of residence. Acceptance is based on 10th grade GPA if the number of applicants exceeds the number of available slots.⁵

3.1.2 Random assignment to classes

When children enroll at elementary school, they are randomly assigned to classes within their district school, which they typically attend through 7th grade. When starting middle school in 8th grade, most students transfer to another school, and are randomly assigned to new classes, which they generally attend through 10th grade.⁶ Throughout the mandatory schooling levels, teaching activities are conducted within the assigned classes.

When assigning students to classes in 1st and 8th grade, the school administration follows guidelines set by the Ministry of Education (the Ministry of Education and Research, 1998, 2017). The ruling principle in class assignment is to create a community of students from different backgrounds, including students in need of special education. As a result, there is no ability tracking in Norwegian elementary and middle schools.

3.1.3 Middle school math teachers

In contrast to elementary school teachers, middle school teachers are usually subject-specific teachers, teaching in only one or two subjects. At middle school, from 8th to 10th grade, students should receive at least 313 hours, á 60 minutes, of education in math (The Norwegian Directorate for Education and Training, 2013). This means that if the math teacher teaches no other subjects, 313 hours is the maximum amount of time students could have spent with this teacher. However, if the students had this math teacher in 10th grade only, they would have been exposed to this particular teacher for about 100 hours, constituting the minimum exposure (disregarding absence from school).

4 Empirical strategy

When investigating the effect of teacher mindset on student outcomes, we follow the standard approach in the literature, and use an educational production function where

⁵Some counties apply the same residential allocation rule as elementary and middle schools, where students are assigned to the nearest school offering their desired educational program. Some other counties allow students to also apply to schools outside the county.

⁶Some places have integrated schools which students attend from 1st to 10th grade. About 23 percent of students in Norway attend integrated schools. These schools tend to be located in rural areas with fewer students (Statistics Norway, 2017b).

we include school fixed effects:

$$(1) \quad y_{ics} = \alpha_s + \beta \text{TM}_{cs} + \delta_1 X_{ics} + \epsilon_{ics}$$

where y_{ics} is the outcome for student i in class c in middle school s . Middle school fixed effects are represented by α_s . TM_{cs} represents the math teacher's mindset, and β our coefficient of interest. X_{ics} represents a vector for control variables on individual and parental characteristics, including gender, birth month, number of siblings, immigrant status, and parental marital status, education and earnings. ϵ_{ics} is an error term for student i in class c in school s , capturing unexplained variations in the outcome variable. Our identifying assumption, and the reason for using school fixed effects, is that within schools, students are randomly assigned to classes, and thus teachers, in Norwegian elementary and middle schools (see [Section 3](#) for details).

We do not, however, observe the desired treatment variable, TM_{cs} , directly in the data, as observing someone's mindset is impossible. Instead we measure this using students' perceptions of their 10th grade math teacher's mindset. Therefore, since we are investigating effects of an unobservable classroom characteristic (teacher mindset), we follow closely the method proposed in [Chetty et al. \(2011\)](#) for looking at the impact of such characteristics. This means that in our analyses we proxy for teacher mindset using students' perceptions and look at the effect of teacher mindset on student outcomes.

Using students' perceptions, however, has its limitations. First, we are relying on the students to be honest and think about the overall picture when assessing their teacher. This could for instance cause problems if students hold grudges and let that affect their answers. However, we do believe using students' perceptions is better than using alternative approaches, like asking the teachers themselves ([Atkins & Wood, 2002](#)). This is because teachers' self-reported scores on mindset (including beliefs in his/her students' possibilities to develop intelligence and talents) are unlikely to be reliable. Second, how students perceive their teacher could be correlated with several teacher characteristics (e.g. conscientiousness, optimism, ability to ask critical reasoning questions, overall effectiveness as an instructor). Hence, while we try to align our definition of mindset with prior literature, we do acknowledge that the "teacher mindset" in our study might also reflect other teacher characteristics.

If we assume that a student's perception of the teacher's mindset is a reliable measure of that teacher's mindset, it may still be endogenous. The reason being that the student's own mindset or experiences might influence how the student perceives both the teacher and their outcomes. For instance, a student that has a good relationship with the teacher and gets more help than the average student might be more likely to perceive the teacher as having a growth mindset compared to students who have a constrained relationship with their teacher. Alternatively, a student that possesses a growth mindset might be more likely to perceive the teacher as also having a growth mindset compared to a student who possesses a fixed mindset. Since getting more help and having a growth mindset are both likely to have positive effects on school performance, student outcomes and their perception of their math teacher's mindset may be correlated. This is referred to as an

"own-observation problem" in the literature (Chetty et al., 2011). To address this, we follow the intuition in Angrist et al. (1999) and construct our treatment variable as the leave-out-mean perceived teacher mindset within the class:

$$(2) \quad \overline{\text{TM}}_{cs}^{-i} = \frac{1}{N-1} \sum_{j=1, j \neq i}^N \text{TM}_{jcs}$$

where $\overline{\text{TM}}_{cs}^{-i}$ is the mean perceived teacher mindset of all students j in class c in middle school s (the TM_{jcs} 's), excluding student i 's own observation ($j \neq i$). However, since we include school fixed effects in our analyses, the treatment variable needs to be adjusted accordingly. We do this using the same approach as Chetty et al. (2011), with the difference between the students' classroom and school peers' perceptions of their math teacher's mindset:

$$(3) \quad \Delta \overline{\text{TM}}_{cs}^{-i} \equiv \overline{\text{TM}}_{cs}^{-i} - \overline{\text{TM}}_s^{-i} = \frac{1}{N-1} \sum_{j=1, j \neq i}^N \text{TM}_{jcs} - \frac{1}{NC-1} \sum_{c=1}^C \sum_{j=1, j \neq i}^N \text{TM}_{jcs}$$

for N students per class and C classes per school.⁷ The leave-out-mean estimator, $\Delta \overline{\text{TM}}_{cs}^{-i}$, serves as our treatment variable, and answers the question: "How much of a growth mindset did your 10th grade math teacher have compared to other math teachers in your school?"

Finally, taking into account possible peer effects on student outcomes, we also include controls for classroom peers' characteristics. Controls for peer characteristics ($\Delta \overline{X}_{cs}^{-i}$) are consistent with the student's individual and parental characteristics and constructed the same way as the treatment variable ($\Delta \overline{\text{TM}}_{cs}^{-i}$). Hence, our preferred model specification when investigating the effects of teacher mindset is:

$$(4) \quad y_{ics} = \alpha_s + \beta \Delta \overline{\text{TM}}_{cs}^{-i} + \delta_1 X_{ics} + \delta_2 \Delta \overline{X}_{cs}^{-i} + \epsilon_{ics}$$

where student i 's treatment variable, $\Delta \overline{\text{TM}}_{cs}^{-i}$, is the difference between his/her classroom and school peers' perceptions of their math teacher's mindset. Through the rest of the paper we refer to this measure as the "teacher mindset" or "treatment" variable.

5 Data

We use data from three different sources. First, data from a survey conducted among students at high schools in two out of the nineteen counties in Norway. We conducted the survey during the first semester of high school, and asked the students a variety of

⁷Fixed values for students per class (N) and classes per school (C) for presentation purposes only. In our regressions, N and C vary across classes and schools.

questions intended to measure their mindset, including educational attitudes and expectations, and perceptions of their 10th grade math teacher's mindset. Survey questions were based on previously used and validated measures from the mindset literature (Burnette et al., 2013; Yeager, Romero, et al., 2016; Bettinger et al., 2018). Second, we use data from administrative records collected and maintained by the county authorities. These records allow us to identify the 10th grade class for each student, and consequently the student's peers in 10th grade. Furthermore, data on students' school performance in middle and high school are also included in these records. Third, registry data from Statistics Norway provides us with individual and parental background characteristics.

A total of 11 072 students consented and participated in the survey (91 percent). Among these, we exclude students with special needs (419), and students in very small (<5) high school classes (550). Another 29 students were excluded from our sample due to technical problems during the survey. Furthermore, we exclude students with missing data on 10th grade GPA (161), math grades (153), or class/school identifier (191). In addition, we exclude students not following regular school progression (251), students in 10th grade classes with less than 10 students (431), and students who did not answer the questions on teacher mindset (2). This leaves us with an analytic sample of 8 885 students; 4 349 boys and 4 536 girls.

5.1 Variable definitions

5.1.1 Treatment variable

In order to construct our treatment variable, teacher mindset, we use the students' perceptions of their 10th grade math teacher's mindset as a proxy for that teacher's mindset, explained in detail in [Section 4](#). The students' perceptions of the teacher's mindset are measured by asking them to think back to their 10th grade math class and answer to which degree, on a six point scale ranging from "Strongly disagree" (1) to "Strongly agree" (6), they agree with the following statements; (1) "If there was something we did not understand, the teacher explained it in a different way" (Explanation); (2) "My math teacher believed that everyone could be good at math" (Think); and (3) "It seemed like our math teacher liked students who were good at math the best" (Like). We invert all answers in order to have high values reflecting a growth mindset. For each student, we calculated the mean of the three (standardized) answers, and re-standardized to create a measure of how each student perceived their 10th grade math teacher's mindset.

5.1.2 Outcome variables

The outcomes of interest, based on our conceptual framework illustrated in [Figure 1](#), include students' general mindset and in particular, their mindset when it comes to math. In addition, we look at students' educational attitudes and expectations, their challenge-seeking behavior, their performance in middle and high school, and their choices regarding track and course selection in high school. Below we address each of these in turn.

Students' mindset: The metric of students' mindset that we focus most extensively on is generated from the following three statements; (1) "You have a certain amount of intelligence, and you really can't do much to change it"; (2) "Your intelligence is something about you that you can't change very much"; and (3) "Being a 'math person' or not is something that you really can't change. Some people are good at math and other people aren't" (Math Measure). The students reported to which degree, on the same six-point scale as for the teacher mindset questions, they agreed with the statements. We invert the scales in order to have high values reflecting a growth mindset. The first two measures have a correlation near 0.70, while the "Math Measure" has correlations of 0.34 and 0.37 with the first and second mindset measures, respectively. We construct a measure of general mindset by taking the mean of the two (standardized) general mindset measures and re-standardize. Correspondingly, the math mindset measure is constructed by standardizing the students' answers on the "Math Measure", and represents a key metric when investigating math performance and engagement.

Students' educational attitudes and expectations: We investigate five different outcomes when looking at attitudes and expectations: (1) if the student would attribute a bad test score to not being good in math ("Attribution"); (2) if the student believes that using another strategy could have increased the score ("Strategy"); (3) if becoming anxious and insecure when thinking about math ("Anxious"); (4) if the student believes high school math will be interesting ("Interest"); and (5) the student's expectation about own performance in math in high school ("Expectations"). Answers on each of the survey questions are standardized with mean zero and standard deviation one. The exact questions used are specified in [Appendix A](#).

Students' challenge-seeking behavior: We measure students' challenge-seeking behavior in two ways. First, we present a hypothetical scenario, where the students are asked to choose between an easy and a more challenging assignment, and we construct an indicator for choosing the challenging one. Second, we ask students to create their own worksheet, consisting of easy, medium and/or challenging tasks.⁸ Subsequently, we construct two variables, one counting the number of challenging and one counting the number of easy tasks in their worksheet, and standardize in order to make them comparable. The exact setting and questions used are specified in [Appendix B](#).

Students' achievement in middle school: Students' achievement in middle school is constructed by standardizing the 10th grade teacher-set math grade. However, this grade is set by the teacher we are investigating the effect of, which could cause problems if teachers with a growth mindset have a different grade practice. Therefore, we also include achievement on the externally graded exam for the one third of students drawn for the math exam at the end of 10th grade.

Students' choices and achievement in high school: First, we construct an indicator for high school track, equal to one for students attending the vocational track and zero for

⁸Afterwards, students would have to solve some randomly selected tasks from their worksheet.

students attending the academic track. Second, we construct an indicator for participating in the advanced math course, equal to one for students who participate and zero for others. Achievement in high school is measured by the (standardized) teacher-set math grade, and indicators for passing first year in high school and passing first-year math.

5.1.3 Control variables

Control variables include indicators for gender (female), immigrant, married mother and parental education (high school, college and advanced college), categorical variable for age (birth month), number of siblings and birth order, and parental income (linear, quadratic and cubic controls). We also include the same controls for peers' individual and parental characteristics. Finally, we include indicators for missing observations on control variables. See [Appendix C](#) for specifics about how missing observations are handled.⁹ All included variables in the different specifications are specified in the full regression on math mindset in [Online Appendix A](#).

5.2 Descriptive statistics

Descriptive statistics for our outcome variables are presented in [Table 1](#). Column 1 of [Table 1](#) presents means for all students. Columns 2 and 3 present means for boys and girls, respectively, and column 4 the p -value for the difference between boys and girls. The descriptive statistics show that among all students, 59 percent report they would choose the challenging one if given the choice between an easy and a more challenging assignment. We find that 32 percent of students are attending the vocational track and 33 percent participate in the advanced math course. The advanced math course is, however, only available to academic track students, and among these, 48 percent chose to participate. 93 percent of the students passed the first year of high school. Remaining variables are standardized with mean zero and standard deviation one.

When comparing across gender, we find that boys are less likely to have a growth mindset in general, but regarding math mindset we do not find any gender difference. Looking at educational attitudes and expectations we find that boys are less likely than girls to attribute a bad score on a math test to being bad in math, and they are more likely to believe that a different strategy might help. Boys are also less likely to be insecure and anxious when thinking about math, they have higher expectations about how they will perform in high school math, and they are somewhat more likely to find math interesting. The next set of outcomes is challenge-seeking behavior, and here we find that boys tend to be more likely to seek challenge: first, 63 percent of the boys would have chosen the challenging assignment, compared to 55 percent of the girls. Second, compared to girls, boys are more likely to choose more of the challenging and less of the easy tasks when given the opportunity to make their own worksheet.

⁹We also include an indicator for assigned group, since data are originally from a randomized controlled trial (RCT).

We also find gender differences regarding real life outcomes, such as choices and achievement. We find that boys tend to do worse than girls in math (10th grade and high school, externally and teacher-set grades). Furthermore, we find that boys (38 percent) are more likely than girls (27 percent) to choose the vocational track. Girls also tend to be slightly more likely than boys to participate in the advanced math course (31 vs 34 percent). However, when restricting the sample to only students attending the academic track, we find that boys are over-represented among those choosing the advanced math course (50 vs 47 percent). There are no gender differences in passing first year in high school.

Descriptive statistics for student, parental and school characteristics are presented in [Table 2](#), with column 1 presenting the statistics for all students, and columns 2 and 3 presenting the statistics separately for boys and girls. In column 4 we show the p -value for the difference in means between boys and girls.

We see that our sample is well balanced across gender. 24 percent of the students are immigrants, the average student has two siblings and is born as the second child in the family. Parents earn on average 1.2 MNOK,¹⁰ and 68 percent of the mothers are married. A total of 14 (15) percent of mothers (fathers) have not completed high school, and 51 (41) percent of mothers (fathers) have a university degree. On average there are 78 students and 4 classes within each school, and 19 students in each class. Finally, columns 2 to 4 show that differences across gender are negligible.

6 Results

6.1 Test of identifying assumption

Our identifying assumption is that, within schools, students are randomly assigned to classes, and thus teachers. To test this assumption, we perform a balance test where we regress the treatment variable on observable student and parental characteristics in three model specifications. The coefficients are reported in [Table 3](#). Each coefficient represents a separate regression. In column 1 we report the coefficients using only school fixed effects, and standard errors clustered at 10th grade school. The treatment variable measures the difference between the classroom and school peers' perceptions of their math teacher's mindset (the $\Delta \overline{\text{TM}}_{cs}^{-i}$). Among all characteristics, we find that only parental earnings is slightly imbalanced across treatment. In columns 2 and 3, the latter being our preferred model, we add controls for individual and parental characteristics, and classroom peers' characteristics, respectively. The sample appears balanced across treatment. The last row provides the p -value for a test for joint significance on all observable characteristics. We find no jointly significant differences for any model specifications. The estimated coefficients in [Table 3](#) provide support for our identifying assumption. In [Appendix D](#) we

¹⁰Corresponding to approximately USD 120 000.

include tests of the identifying assumption across gender, and find no significant differences.

6.2 Estimated effects

In line with our conceptual framework, illustrated in [Figure 1](#), the outcomes of interest are students' mindset, their educational attitudes and expectations, their challenge-seeking behavior, academic choices and achievement. All outcomes are described in detail in [Section 5](#).

6.2.1 Students' mindset

In columns 1 to 3 of [Table 4](#) we present the estimated effects on students' general mindset, and in columns 4 to 6 we present the estimated effects on students' mindset when it comes to math. In all specifications, we include fixed effects for middle school. The treatment variable is the difference between classroom and school peers' perceptions of their teacher's mindset ($\Delta \overline{\text{TM}}_{cs}^{-i}$). In columns 2 and 5 we include controls for individual and parental characteristics, before finally in columns 3 and 6 we add controls for classroom peers' individual and parental characteristics, our preferred model specification. With regard to students' general mindset, we do not find any significant effects. The estimated coefficients for girls are, however, larger than for boys in all of the specifications, but not significantly different.

Students' math mindset, on the other hand, seems to be affected by the 10th grade math teacher's mindset. In our simplest specification, including only school fixed effects, the effect sizes are 2.9 percent of a standard deviation for the entire sample. They stay significant with an effect size of 2.7 percent of a standard deviation for the full sample when adding controls for individual and parental characteristics, as well as classroom peers' individual and parental characteristics. The corresponding estimates when dividing the sample based on gender are 2.1 (not significant) for boys and 3.8 for girls. This implies that having a teacher with a one standard deviation more of a growth mindset than the average teacher in the school increases students' mindset by 2.7 percent of a standard deviation (3.8 for girls). However, none of the coefficients prove to be significantly different across gender.¹¹

6.2.2 Students' educational attitudes and expectations

[Table 5](#) presents the estimated coefficients on students' educational attitudes and expectations. We report estimates only from our preferred model specification (as presented in columns 3 and 6 of [Table 4](#)). Treatment effects will be referred to as effects of having

¹¹To facilitate interpretation of the effect sizes, we replicated the analyses replacing the continuous treatment variable with an indicator for having a growth mindset teacher. The corresponding estimates are 5.1 and 8.3 percent of a standard deviation, respectively. Analyses available upon request.

a teacher with more of a growth mindset.¹² Column 1 presents the estimated coefficients with regards to attribution. We find that having a teacher with more of a growth mindset has a negative effect (3.3 percent) on attributing a bad test score to "not having what it takes to succeed in math". The effect is similar across gender (3.1 and 3.6 for boys and girls, respectively).

In column 2 we present the estimated effects on how likely students are to believe that using a different strategy would have helped on a bad test score. Here we find a positive effect of having a teacher with more of a growth mindset of 3 percent. The effect appears to be more prevalent among boys, where we find a significant effect of 4.3 percent of a standard deviation (versus 1.5, and not significant, for girls).

Estimated coefficients on students becoming anxious and insecure when thinking about math in high school are presented in column 3. We find that having a teacher with more of a growth mindset has a negative effect of 4.5 percent of a standard deviation on students feeling anxious and insecure. The corresponding coefficients for boys and girls are 5.2 and 4.2, respectively.

Column 4 presents results for students' expectations about how they will perform in high school math. Having a teacher with more of a growth mindset has a positive effect on students' expectations about own performance, with estimated effects of 3.9 percent of a standard deviation; 4.7 for boys and 3.2 for girls when dividing the sample based on gender.

In column 5 we report results for our last outcome of interest with regards to attitudes and expectations. On students' interest in high school math we find a positive effect of 3.7 percent of a standard deviation. A breakdown by gender shows 4.1 and 3.6 for boys and girls, respectively.

Due to concerns about multiple hypothesis testing, we ran a test of joint significance, reported below each sample. Estimated coefficients are jointly significant for the full sample and the subsample of boys. The subsample of girls, however, shows a joint significance p-value below the conventional level.

Consistent with results reported above, the p-value reported in the last line of the table suggests that none of the estimated coefficients are significantly different for boys and girls. This suggests that there are no differences in terms of educational attitudes and expectations of how boys and girls are affected by their math teacher's mindset.

¹²Notably, the treatment variable measures differences in perceived mindset between teachers at a given school: "How much of a growth mindset did your 10th grade math teacher have compared to other math teachers in your school?"

6.2.3 Students' challenge-seeking behavior

Estimated effects of teacher mindset on students' challenge-seeking behavior are reported in [Table 6](#). Column 1 presents estimates for choosing the challenging assignment if given a choice between a challenging and an easier assignment. Columns 2 and 3 report the estimates for number of challenging and easy tasks from the "Make your own worksheet" assignment. We find a positive effect of 1.1 percent on choosing the challenging assignment in column 1 (marginally significant). The result seems to be entirely driven by boys (2.3 percent), and the p-value reported in the last line of the table confirms that the estimated coefficients for boys and girls are significantly different. For the "Make-your-own-worksheet" assignment we find indications that students who have a teacher with more of a growth mindset choose more of the challenging and less of the easy tasks, but the estimated coefficients are small and not statistically significant.

6.2.4 Students' achievement in middle school

[Table 7](#) reports the estimated coefficients with regards to math performance in 10th grade. In column 1 we report the estimated effects on the teacher-set math grades for the full sample, and in column 2 the corresponding effects on the subsample of students drawn for the math exam. In column 3 we report the estimated effects on the externally graded exam. In column 1 we see that being exposed to a math teacher with more of a growth mindset has a positive effect on math grades (3.1 percent), with no differences across gender. Notably, this grade is set by the teacher we are investigating the effect of, so if teachers with more of a growth mindset tend to be more lenient when setting grades, this estimate is upward biased. However, calculations show that the correlation between grade-inflated teachers and teachers with more of a growth mindset is below 5 percent, suggesting this is not affecting our estimates, ref. [Appendix E](#).

The estimated effects on the math exam grade in column 3 are small and statistically insignificant. Notably, this may be due to the smaller sample size and loss of precision, supported by similar results on the teacher-set grade on the same smaller sample in column 2. The similarity of results in columns 2 and 3 also supports the finding that teachers with more of a growth mindset are not more likely to be grade-inflated than other teachers.

6.2.5 Students' choices and achievement in high school

Our last set of outcomes is students' choices and achievement in high school, with estimated coefficients reported in [Table 8](#). In column 1 we report effects on the choice between vocational and academic track. We do not find any effects with regards to this choice in either of our samples. In column 2 we investigate the effect on participation in the advanced math course, and find a significant and positive effect of 0.9 percent, which appears to be driven by girls (1.4 percent). However, the p-value in the lower row of the table suggests that the coefficients for boys and girls are not significantly different.

The advanced math course is only available to students attending the academic track. Since we found no evidence of teacher mindset affecting choice of academic versus vocational track, we replicate the analyses on advanced math on a sample excluding students attending the vocational track, see column 3. As expected, among students attending the academic track, we find a somewhat larger effect than in the full sample, with a significant and positive effect of 1.5 percent. There are no gender differences in this restricted sample, reflecting that boys are over-represented in the vocational track.

A possible consequence of stimulating students to choose the advanced math course, is that some of these students are seeking more challenge than they are capable of handling. In that case, this may affect both the grade and the pass rate. In column 4 we investigate the impact on passing first year in high school for the full sample of students, and in column 5 we restrict the sample to students attending the academic track. Results indicate that being exposed to a growth mindset teacher has a negative effect on passing, but the estimates are all small or insignificant. The effect on math grades in high school is presented in columns 6 and 7 of Table 8, with the full sample of students in column 6, and a sample restricted to students attending the academic track in column 7. Among all students in our sample, the coefficient on math grade is close to zero, and not significant. For the restricted sample, we find a negative effect of around 2 percent of a standard deviation (also not significant).

Table 8 does not fully address the concern about students taking on more challenge than they can handle. In particular, this is relevant for students participating in the advanced math course. However, since choice of math course is affected by teacher mindset, restricting the sample to include only these students will make the sample endogenous. It is therefore challenging to separate the effect on choice of math course from the effect on passing and math grades. Below we deal with this by estimating the upper and lower bounds of the effects on passing math and math grades in the first year of high school. The sample is restricted to students attending the academic track and participating in the advanced math course.

6.3 Bounds

If having a teacher with more of a growth mindset makes students more challenge-seeking (Mueller & Dweck, 1998; Blackwell et al., 2007) and, as a result, more likely to participate in the advanced math course, the estimated coefficients above with regards to math grades and passing the first year of high school do not fully take the influence of the teacher into account. The reason being that the treatment makes students choose a path where passing and high grades might be harder to achieve. Therefore we follow the intuition in Angrist et al. (2006), and estimate lower and upper bounds for math grade and passing math the first year of high school for students choosing the advanced math course.

We estimate the lower bound by restricting the sample to students attending the academic track and participating in the advanced math course, and perform the analyses

on this restricted sample. In order to identify the upper bounds of the true treatment estimate, we exclude from this sample students who participate in the advanced math course as a response to the influence from their teacher’s mindset. We identify these students by first estimating the proportion of students whose choice of math course is affected. Second, we assume that the affected students are those at the lower end of the grade distribution who also had a teacher with a growth mindset. We estimate the proportion of students whose choice of math course is affected using the following regression

$$(5) \quad y_{ics} = \alpha_s + \beta \text{GM} + \delta_1 X_{ics} + \delta_2 \Delta \overline{X}_{cs}^{-i} + \epsilon_{ics}$$

which is similar to Equation (4) with the exception of the treatment variable. Rather than utilizing the continuous variable, $\Delta \overline{TM}_{cs}^{-i}$, we construct an indicator, GM, for teachers having a growth mindset. The indicator takes the value one if $\Delta \overline{TM}_{cs}^{-i}$ is larger than zero, and zero if not. This implies labeling about half of the teachers as having a growth mindset and half of them as having a fixed mindset.

The coefficient β represents the proportion of students who participate in the advanced math course as a result of being influenced by their growth mindset teacher. This is the proportion of students we want to exclude from our sample when calculating the upper bounds. We drop the fraction of students corresponding to β in the lower end of the grade distribution among students who participate in the advanced math course and had a teacher with a growth mindset. When analyzing treatment effects on this smaller sample, the effect sizes will represent an upper bound of the true treatment effect. The outcomes of interest when identifying lower and upper bounds are math grade and passing math in the first year of high school. Estimated coefficients are presented in Table 9.

In column 1 we estimate the proportion of students whose choice of course is affected (advanced math).¹³ We find that 4.8 percent of students on the academic track participate in the advanced math course due to having a teacher with a growth mindset. The corresponding coefficients for boys and girls are 5.9 and 3.8 percent. All estimates is statistically significant.

When estimating the lower bounds in columns 2 and 4, we restrict the sample further to include only students participating in the advanced math course. The estimated coefficients in columns 2 and 4 represent the lower bound of the effect of having a teacher with more of a growth mindset. The estimates suggest that there are no effects on passing math, but that there is a small negative effect of about 3 percent of a standard deviation on math grades (marginally significant for the full sample, and not significant for the subsamples with regards to gender).

In columns 3 and 5 we exclude 4.8 percent of the lowest achieving students (5.9 percent of boys and 3.8 percent of girls) among those having a teacher with a growth

¹³That math grade was missing for 283 students, who were therefore excluded from these analyses.

mindset,¹⁴ and argue that these are the students who participate in the advanced math course because of their teacher. This means that we exclude 69 students when performing the analyses on upper bounds. The estimated upper bound effects are reported in column 3 (passing math), and in column 5 (math grade). We find a positive and significant effect of 0.3 percent on passing math, and the effects seem to be driven by boys (0.5 percent). On the math grades, the estimates are positive or very close to zero, but statistically insignificant.

Our result when looking at the lower bound indicates a negative treatment effect on math grades for the subsample of students participating in the advanced math course. The effect seems to be driven by students participating in the advanced course because of having a teacher with a growth mindset, with upper bound effects close to zero (or positive). However, we find no indications that they are worse off in terms of passing math, and among the students who would have chosen the advanced course anyway (the upper bound), we actually find a (small) positive effect on passing math in the first year of high school.

6.4 Threats to validity

6.4.1 Attenuation bias

Even though we remove the upward bias due to the own-observation problem using a leave-out-mean estimator ($\Delta \overline{\text{TM}}_{cs}^{-i}$), our estimated coefficients will still be attenuated relative to the true coefficients because the perceived teacher mindset is a noisy measure of the teacher's "true" mindset (Chetty et al., 2011). We follow the approach in Appendix B of Chetty et al. (2011) and calculate the attenuation bias in our estimated coefficients at 2.6 percent (for calculations see [Appendix F](#)). This means that our estimated effects will be attenuated relative to the true effect by 2.6 percent.

6.4.2 Peer effects

In the analysis presented above, we have ignored the variation in perceived teacher mindset due to peer effect. For instance, in their study investigating the effect of class quality on test scores and earnings, Chetty et al. (2011) expect a higher ability student to have a positive impact on the class, thereby violating the assumption that class quality is orthogonal to the individual error term with regards to intrinsic academic ability. Translated to our context, this implies that a student may affect how his or her peers perceive their teacher's mindset through how the student talks about the teacher, possibly reflecting his or her own mindset. This leads to a bias in the estimated coefficient on teacher mindset, due to the fact that some of the perceived teacher mindset comes from the teacher's behavior (labeled the pure teacher effect by Chetty et al. (2011)) and some of

¹⁴Corresponding to the effect sizes in column 1.

it comes from the student's peers (and their perception of the teacher), thus introducing reflection bias to the estimated coefficients.

Given that we only observe the student's perception at one single point in time, we have limited possibilities to control for the potential influence a student might have on his or her peers' perceptions (Chetty et al., 2011; Kraft, 2019). This means that even if we utilize a leave-out-mean approach, the student might still influence the treatment variable, $\Delta \overline{\text{TM}}_{cs}^{-i}$, through his or her impact on peers. However, as Duckworth (2017) points out about questions like these; the teacher mindset questions that students are asked in the survey represent thoughts that if asked, people usually have an answer, but might not be aware of their thoughts on the matter until the question is raised. It is therefore unlikely that students have discussed the topic with their peers before being asked about it in the survey, limiting the potential peer influence on their answers. Unfortunately, we are not able to completely remove this concern.

7 Conclusion

Students' mindset, their educational attitudes and expectations, the way they respond to challenge and setbacks as well as their performance might be affected by their teachers. In this paper we investigate the effect of having a math teacher with more of a growth mindset than other math teachers in the school in 10th grade, i.e. the last year in middle school, on outcomes in the first year of high school. The focus of this paper is on girls in STEM, thus our breakdowns are based on gender and focus especially on outcomes related to math. We try to align our treatment variable with prior research and describe it as having a teacher with more of a growth mindset. However, we are using students' perceptions of their teacher's mindset to construct this variable, which may also reflect other positive teacher characteristics (e.g. conscientiousness, optimism, ability to ask critical reasoning questions, overall effectiveness as an instructor).

When asking students about their reactions to getting a bad score on a math test, their anxiety level when it comes to math, their expectations about own performance and their interest in the subject, our descriptive statistics reveal answers in line with previous empirical findings (Petersen & Hyde, 2014; Eccles et al., 1983). Boys seem to be less prone to thinking that a bad test score labels them as not being good in math, they are less anxious, have higher expectations about own performance and show more interest in the subject. Previous literature suggests that attitudes like these might be one of the reasons for different STEM participation between boys and girls (Bandura et al., 2001; Eccles, 1994; Petersen & Hyde, 2014). Students' ability when it comes to math, however, does not seem to coincide with the expectations and interest in the subject, as we only find negligible differences between boys and girls with regards to achievement in math, also in line with previous findings (Petersen & Hyde, 2014; Lindberg et al., 2010; Hyde et al., 2008).

Our estimated effects reveal that being exposed to a teacher with more of a growth mindset in 10th grade, when important high school decisions and consequently future career choices are being made, is beneficial for students' mindset when it comes to math. Our estimates also show that having a teacher with more of a growth mindset has positive effects on the other aspects as well, with students being less likely to attribute a bad score on a math test to not being that good at math, and more likely to believe that a different strategy might have helped. In addition, they are less anxious, have higher expectations about own performance and show more interest in the subject. Despite initial beliefs, that having a teacher with more of a growth mindset is more important for girls, we do not find evidence of different effects across student gender in terms of mindset or attitudes.

With regards to the focus of our paper - girls and STEM - the best indication we have is the choice of track and participation in the advanced math course in high school. The reason being that these choices are crucial to continuing towards a STEM degree in higher education, and thus a STEM career later on. In our analyses we do not find any effect on the choice of track, but we do find a positive effect on participation in the more advanced math course. Furthermore, this effect seems to be similar for both genders. This is an important finding as several studies have highlighted participation in the advanced math course as a critical gateway into the ever more important STEM fields for students (Sells, 1980; Ma & Johnson, 2008; Petersen & Hyde, 2014; Freeman & Aspray, 1999).

When looking at challenge-seeking behavior and attitudes in general, as well as achievement in middle school, we find small positive effects of having a teacher with more of a growth mindset. The effect on challenge-seeking attitudes, however, seems to be driven primarily by boys. When investigating achievement in high school, we find some negative effects on passing and math grades. One explanation for this could be, as supported by our findings and pointed out in the literature, effects on challenge-seeking behavior (Mueller & Dweck, 1998; Blackwell et al., 2007; Yeager, Romero, et al., 2016; Bettinger et al., 2018). In other words, students who have a teacher with more of a growth mindset opt for more challenge, and as a result get lower grades.

We investigate this further by calculating the upper and lower bounds among students who participated in the most challenging math course in high school, and find a negative effect among all students. However, when excluding students who chose the advanced math course as a response to their 10th grade math teacher's mindset, we find indications of positive effects on math grades in high school as well (but not significant). With regards to the students opting for more challenge, and being worse off in terms of math grades, one important question is; Does this choice affect their pass rate? This does not seem to be the case when looking at the pass rate for math in the first year of high school, with the effect being close to zero on the lower bound. For students who would have chosen the advanced math course anyway (the upper bound), there actually seems to be a positive effect on the pass rate, particularly for boys. The findings here beg the question; What is most beneficial for students; getting a better grade in a lower level course or being exposed to a more advanced curriculum?

With positive effects on participation in the more advanced math course and on students' mindset and attitudes towards math, our findings suggest that having a teacher with more of a growth mindset supports students' belief and confidence. However, more research is needed in order to understand gender specific effects. In particular, it is important to understand what motivates girls to advance within the STEM field. The complexity of the future labor market requires girls also to work within this field (Petersen & Hyde, 2014; Freeman & Aspray, 1999).

Future research should also incorporate other aspects of the school, not just the math teacher, like the overall culture or the group of teachers that students are exposed to. The effects of having a math teacher with a growth mindset is likely to interact with the mindset of the general teacher community at the school: a math teacher with a growth mindset in a fixed mindset teacher community may have a harder job in convincing students to believe in their abilities to learn, than if surrounded by colleagues with a growth mindset.

References

- Alan, S., & Ertac, S. (2017). Belief in Hard Work and Prosocial Behavior: Evidence from a Randomized Field Experiment. Retrieved from http://conference.iza.org/conference_files/Cognitive_2017/alan_s1276.pdf
- Andersen, S. C., & Nielsen, H. S. (2016). Reading intervention with a growth mindset approach improves children's skills. *Proceedings of the National Academy of Sciences*, *113*(43), 12111–12113. doi: <https://doi.org/10.1073/pnas.1607946113>
- Angrist, J. D., Bettinger, E., & Kremer, M. (2006). Long-Term Educational Consequences of Secondary School Vouchers: Evidence from Administrative Records in Colombia. *American Economic Review*, *96*(3), 847–862. doi: <https://doi.org/10.1257/aer.96.3.847>
- Angrist, J. D., Imbens, G. W., & Krueger, A. B. (1999). Jackknife instrumental variables estimation. *Journal of Applied Econometrics*, *14*(1), 57–67. doi: [https://doi.org/10.1002/\(SICI\)1099-1255\(199901/02\)14:1<57::AID-JAE501>3.0.CO;2-G](https://doi.org/10.1002/(SICI)1099-1255(199901/02)14:1<57::AID-JAE501>3.0.CO;2-G)
- Araujo, M. C., Carneiro, P., Cruz-Aguayo, Y., & Schady, N. (2016). Teacher Quality and Learning Outcomes in Kindergarten. *The Quarterly Journal of Economics*, *131*(3), 1415–1453. doi: <https://doi.org/10.1093/qje/qjw016>
- Aronson, J., Fried, C. B., & Good, C. (2002). Reducing the Effects of Stereotype Threat on African American College Students by Shaping Theories of Intelligence. *Journal of Experimental Social Psychology*, *38*(2), 113–125. doi: <https://doi.org/10.1006/jesp.2001.1491>
- Atkins, P. W., & Wood, R. E. (2002). Self-versus others' ratings as predictors of assessment center ratings: Validation evidence for 360-degree feedback programs. *Personnel Psychology*, *55*(4), 871–904. doi: <https://doi.org/10.1111/j.1744-6570.2002.tb00133.x>
- Bandura, A., Barbaranelli, C., Caprara, G. V., & Pastorelli, C. (2001). Self-Efficacy Beliefs as Shapers of Children's Aspirations and Career Trajectories. *Child Development*, *72*(1), 187–206. doi: <https://doi.org/10.1111/1467-8624.00273>
- Beilock, S. L., Gunderson, E. A., Ramirez, G., & Levine, S. C. (2010). Female teachers' math anxiety affects girls' math achievement. *Proceedings of the National Academy of Sciences*, *107*(5), 1860–1863. doi: <https://doi.org/10.1073/pnas.0910967107>
- Bettinger, E. P., & Long, B. T. (2005). Do Faculty Serve as Role Models? The Impact of Instructor Gender on Female Students. *American Economic Review*, *95*(2), 152–157. doi: <https://doi.org/10.1257/000282805774670149>
- Bettinger, E. P., Ludvigsen, S., Rege, M., Solli, I. F., & Yeager, D. (2018). Increasing perseverance in math: Evidence from a field experiment in Norway. *Journal of Economic Behavior & Organization*, *146*, 1–15. doi: <https://doi.org/10.1016/j.jebo.2017.11.032>
- Blackwell, L. S., Trzesniewski, K. H., & Dweck, C. S. (2007). Implicit Theories of Intelligence Predict Achievement Across an Adolescent Transition: A Longitudinal Study and an Intervention. *Child Development*, *78*(1), 246–263. doi: <https://doi.org/10.1111/j.1467-8624.2007.00995.x>
- Blazar, D., & Kraft, M. A. (2017). Teacher and Teaching Effects on Students' Attitudes and Behaviors. *Educational evaluation and policy analysis*, *39*(1), 146–170. doi: <https://doi.org/10.3102/0162373716670260>

- Brooks, R., & Goldstein, S. (2008). The Mindset of Teachers Capable of Fostering Resilience in Students. *Canadian Journal of School Psychology, 23*(1), 114–126. doi: <https://doi.org/10.1177/0829573508316597>
- Burnette, J. L., O'boyle, E. H., VanEpps, E. M., Pollack, J. M., & Finkel, E. J. (2013). Mind-sets matter: A meta-analytic review of implicit theories and self-regulation. *Psychological Bulletin, 139*(3), 655–701. doi: <https://doi.org/10.1037/a0029531>
- Card, D., & Payne, A. A. (2017). High school choices and the gender gap in STEM. *NBER Working Paper No. 23769*. doi: <https://doi.org/10.3386/w23769>
- Carrell, S. E., Page, M. E., & West, J. E. (2010). Sex and Science: How Professor Gender Perpetuates the Gender Gap. *The Quarterly Journal of Economics, 125*(3), 1101–1144. doi: <https://doi.org/10.1162/qjec.2010.125.3.1101>
- Chetty, R., Friedman, J., Hilger, N., Saez, E., Schanzenbach, D., & Yagan, D. (2011). How Does Your Kindergarten Classroom Affect Your Earnings? Evidence from Project Star. *The Quarterly Journal of Economics, 126*(4), 1593–1660. doi: <https://doi.org/10.1093/qje/qjr041>
- Chetty, R., Friedman, J. N., & Rockoff, J. E. (2014a). Measuring the Impacts of Teachers I: Evaluating Bias in Teacher Value-Added Estimates. *American Economic Review, 104*(9), 2593–2632. doi: <https://doi.org/10.1257/aer.104.9.2593>
- Chetty, R., Friedman, J. N., & Rockoff, J. E. (2014b). Measuring the Impacts of Teachers II: Teacher Value-Added and Student Outcomes in Adulthood. *American Economic Review, 104*(9), 2633–2679. doi: <https://doi.org/10.1257/aer.104.9.2633>
- Dobronyi, C. R., Oreopoulos, P., & Petronijevic, U. (2019). Goal Setting, Academic Reminders, and College Success: A Large-Scale Field Experiment. *Journal of Research on Educational Effectiveness, 12*(1), 38–66. doi: <https://doi.org/10.1080/19345747.2018.1517849>
- Duckworth, A. L. (2017). *GRIT: Why passion and resilience are the secrets to success*. London: Vermilion, an imprint of Ebury Publishing.
- Dweck, C. S. (1986). Motivational processes affecting learning. *American Psychologist, 41*(10), 1040–1048. doi: <https://doi.org/10.1037/0003-066X.41.10.1040>
- Dweck, C. S. (2006). *Mindset: The New Psychology of Success*. New York: The Random House Publishing Group.
- Dweck, C. S. (2014). Teachers' Mindsets. *Educational Horizons, 93*(2), 10–15. doi: <https://doi.org/10.1177/0013175X14561420>
- Dweck, C. S., & Leggett, E. L. (1988). A social-cognitive approach to motivation and personality. *Psychological Review, 95*(2), 256–273. doi: <https://doi.org/10.1037/0033-295X.95.2.256>
- Eccles, J. S. (1994). Understanding Women's Educational And Occupational Choices: Applying the Eccles et al. Model of Achievement-Related Choices. *Psychology of Women Quarterly, 18*(4), 585–609. doi: <https://doi.org/10.1111/j.1471-6402.1994.tb01049.x>
- Eccles, J. S., Adler, T. F., Futterman, R., Goff, S. B., Kaczala, C. M., & Meece, J. L. (1983). Expectancies, values and academic behaviors. *Achievement and Achievement Motives*.

- Freeman, P., & Aspray, W. (1999). *The Supply of Information Technology Workers in the United States* (Tech. Rep.). ERIC. Retrieved from <https://files.eric.ed.gov/fulltext/ED459346.pdf>
- Good, C., Aronson, J., & Inzlicht, M. (2003). Improving adolescents' standardized test performance: An intervention to reduce the effects of stereotype threat. *Journal of Applied Developmental Psychology, 24*(6), 645–662. doi: <https://doi.org/10.1016/j.appdev.2003.09.002>
- Good, C., Rattan, A., & Dweck, C. S. (2012). Why do women opt out? Sense of belonging and women's representation in mathematics. *Journal of Personality and Social Psychology, 102*(4), 700–717. doi: <https://doi.org/10.1037/a0026659>
- Griffith, A. L. (2010). Persistence of women and minorities in STEM field majors: Is it the school that matters? *Economics of Education Review, 29*(6), 911–922. doi: <https://doi.org/10.1016/j.econedurev.2010.06.010>
- Grossman, P., Loeb, S., Cohen, J., & Wyckoff, J. (2013). Measure for Measure: The Relationship between Measures of Instructional Practice in Middle School English Language Arts and Teachers' Value-Added Scores. *American Journal of Education, 119*(3), 445–470. doi: <https://doi.org/10.1086/669901>
- Hanushek, E. A. (1971). Teacher Characteristics and Gains in Student Achievement: Estimation Using Micro Data. *American Economic Review, 61*(2), 280–288. Retrieved from <https://www.jstor.org/stable/1817003>
- Heckman, J. J., Stixrud, J., & Urzua, S. (2006). The Effects of Cognitive and Noncognitive Abilities on Labor Market Outcomes and Social Behavior. *Journal of Labor Economics, 24*(3), 411–482. doi: <https://doi.org/10.1086/504455>
- Heggart, K. (2017). Early childhood student teachers: It pays to belong to the IEU. *News-month, 37*(7). Retrieved from <https://search.informit.com.au/documentSummary;dn=211234753189676;res=IELHSS>
- Hyde, J. S., Lindberg, S. M., Linn, M. C., Ellis, A. B., & Williams, C. C. (2008). Gender Similarities Characterize Math Performance. *Science, 321*(5888), 494–495. doi: <https://doi.org/10.1126/science.1160364>
- Kahn, S., & Ginther, D. (2017). Women and STEM. *NBER Working Paper No. 23525*. doi: <https://doi.org/10.3386/w23525>
- Kane, T. J., Taylor, E. S., Tyler, J. H., & Wooten, A. L. (2011). Identifying Effective Classroom Practices Using Student Achievement Data. *Journal of Human Resources, 46*(3), 587–613. doi: <https://doi.org/10.3368/jhr.46.3.587>
- Kautz, T., Heckman, J. J., Diris, R., Ter Weel, B., & Borghans, L. (2014). Fostering and Measuring Skills: Improving Cognitive and Non-Cognitive Skills to Promote Lifetime Success. *NBER Working Paper No. 20749*. doi: <https://doi.org/10.3386/w20749>
- Kraft, M. A. (2019). Teacher Effects on Complex Cognitive Skills and Social-Emotional Competencies. *Journal of Human Resources, 54*(1), 1–36. doi: <https://doi.org/10.3368/jhr.54.1.0916.8265R3>
- Lindberg, S. M., Hyde, J. S., Petersen, J. L., & Linn, M. C. (2010). New trends in gender and mathematics performance: A meta-analysis. *Psychological Bulletin, 136*(6), 1123–1135. doi: <https://doi.org/10.1037/a0021276>

- Ma, X., & Johnson, W. (2008). Mathematics as the critical filter: Curricular effects on gendered career choices. In *In H. M. G. Watt & J. S. Eccles (Eds.), Gender and occupational outcomes: Longitudinal assessments of individual, social, and cultural influences* (pp. 55–83). Washington, DC: American Psychological Association. doi: <https://doi.org/10.1037/11706-002>
- Markussen, E., & Gloppen, S. K. (2012). *Påbygg – et gode eller en nødløsning? En studie av påbygging til generell studiekompetanse i Østfold, Akershus, Buskerud, Rogaland og Nord-Trøndelag skoleåret 2010-2011* (Tech. Rep.). Retrieved from <https://nifu.brage.unit.no/nifu-xmlui/bitstream/handle/11250/280848/NIFUrapport2012-2.pdf?sequence=1&isAllowed=y>
- Ministry of Education and Research. (1998). *Lov 17. juli 1998 om grunnskolen og den vidaregåande opplæringa (the Education Act)* [Government Document]. Retrieved from <https://lovdata.no/dokument/ML/lov/1998-07-17-61> (Opplæringslova)
- Ministry of Education and Research. (2017). *Veiledning om organisering av elevene* [Government Document]. Online. Retrieved from https://www.regjeringen.no/contentassets/f94154aa3d2b491ba1ac2f7f658cb019/veiledning-om-organisering-av-elevene_oppdatert-april-2017.pdf (Accessed November 2018)
- Mueller, C. M., & Dweck, C. S. (1998). Praise for Intelligence Can Undermine Children's Motivation and Performance. *Journal of Personality and Social Psychology*, *75*(1), 33–52. doi: <https://doi.org/10.1037/0022-3514.75.1.33>
- Murnane, R. J. (1975). *The Impact of School Resources on the Learning of Inner City Children*. Cambridge: Ballinger Pub. Co.
- Nix, S., Perez-Felkner, L., & Thomas, K. (2015). Perceived mathematical ability under challenge: a longitudinal perspective on sex segregation among STEM degree fields. *Frontiers in Psychology*, *6*, Article 530. doi: <https://doi.org/10.3389/fpsyg.2015.00530>
- Paunesku, D., Walton, G. M., Romero, C., Smith, E. N., Yeager, D. S., & Dweck, C. S. (2015). Mind-Set Interventions Are a Scalable Treatment for Academic Underachievement. *Psychological Science*, *26*(6), 784–793. doi: <https://doi.org/10.1177/0956797615571017>
- Perez-Felkner, L., Nix, S., & Thomas, K. (2017). Gendered Pathways: How Mathematics Ability Beliefs Shape Secondary and Postsecondary Course and Degree Field Choices. *Frontiers in Psychology*, *8*, Article 386. doi: <https://doi.org/10.3389/fpsyg.2017.00386>
- Petersen, J. L., & Hyde, J. S. (2014). Gender-Related Academic and Occupational Interests and Goals. In *Advances in Child Development and Behavior* (Vol. 47, pp. 43–76). US/UK: Academic Press, an imprint of Elsevier. doi: <https://doi.org/10.1016/bs.acdb.2014.04.004>
- Petersen, J. L., & Hyde, J. S. (2017). Trajectories of self-perceived math ability, utility value and interest across middle school as predictors of high school math performance. *Educational Psychology*, *37*(4), 438–456. doi: <https://doi.org/10.1080/01443410.2015.1076765>
- Riegle-Crumb, C., Farkas, G., & Muller, C. (2006). The Role of Gender and Friendship in Advanced Course Taking. *Sociology of Education*, *79*(3), 206–228. doi: <https://doi.org/10.1177/003804070607900302>

- Rockoff, J. E. (2004). The Impact of Individual Teachers on Student Achievement: Evidence from Panel Data. *American Economic Review*, *94*(2), 247–252. doi: <https://doi.org/10.1257/0002828041302244>
- Samuelsson, M., & Samuelsson, J. (2016). Gender differences in boys' and girls' perception of teaching and learning mathematics. *Open Review of Educational Research*, *3*(1), 18–34. doi: <https://doi.org/10.1080/23265507.2015.1127770>
- Sells, L. W. (1980). Mathematics: The Invisible Filter. *Engineering Education*, *70*(4), 340–341. Retrieved from <https://eric.ed.gov/?id=EJ216630>
- Sjoquist, D. L., & Winters, J. V. (2015). State Merit Aid Programs and College Major: A Focus on STEM. *Journal of Labor Economics*, *33*(4), 973–1006. doi: <https://doi.org/10.1086/681108>
- Statistics Norway. (2017a). *Elevar i grunnskolen*. Retrieved from <https://www.ssb.no/utdanning/statistikker/utgrs> (Table 4 accessed November 2018)
- Statistics Norway. (2017b). *Elevar i grunnskolen*. Online. Retrieved from <https://www.ssb.no/utdanning/statistikker/utgrs> (Table 1 accessed November 2018)
- The Norwegian Directorate for Education and Training. (2013). *Læreplan i matematikk fellesfag (mat1-04)*. Online. Retrieved from <http://data.udir.no/k106/MAT1-04.pdf> (Accessed February 2019)
- Xie, Y., & Shauman, K. A. (2005). *Women in Science: Career Processes and Outcomes*. Cambridge: Harvard University Press.
- Yeager, D. S., & Dweck, C. S. (2012). Mindsets That Promote Resilience: When Students Believe That Personal Characteristics Can Be Developed. *Educational Psychologist*, *47*(4), 302–314. doi: <https://doi.org/10.1080/00461520.2012.722805>
- Yeager, D. S., Hanselman, P., Walton, G. M., Murray, J. S., Crosnoe, R., Muller, C., ... others (2019). A national experiment reveals where a growth mindset improves achievement. *Nature*, *573*(7774), 364–369.
- Yeager, D. S., Romero, C., Paunesku, D., Hulleman, C. S., Schneider, B., Hinojosa, C., ... others (2016). Using design thinking to improve psychological interventions: The case of the growth mindset during the transition to high school. *Journal of Educational Psychology*, *108*(3), 374–391. doi: <https://doi.org/10.1037/edu0000098>
- Yeager, D. S., Walton, G. M., Brady, S. T., Akcinar, E. N., Paunesku, D., Keane, L., ... others (2016). Teaching a lay theory before college narrows achievement gaps at scale. *Proceedings of the National Academy of Sciences*, *113*(24), 3341–3348. doi: <https://doi.org/10.1073/pnas.1524360113>

Table 1: Descriptive statistics student outcomes

	(1)	(2)	(3)	(4)
	All students	Boys	Girls	<i>p</i> -value
Mindset measures				
General mindset	0.00	-0.04	0.03	0.001
Math mindset	0.00	0.00	-0.00	0.927
Educational attitude and expectations				
Attribution ('Not so good')	0.00	-0.17	0.16	0.000
New strategy	0.00	0.03	-0.03	0.015
Anxious	0.00	-0.28	0.27	0.000
Expectations	0.00	0.13	-0.12	0.000
Interest	0.00	0.03	-0.03	0.006
Challenge-seeking behavior				
Challenging assignment	0.59	0.63	0.55	0.000
Challenging worksheet	0.00	0.10	-0.09	0.000
Easy worksheet	0.00	-0.06	0.05	0.000
Achievement in middle school				
10th grade math	0.00	-0.07	0.07	0.000
10th grade math exam	0.00	-0.04	0.04	0.014
Choices and achievement in high school				
Vocational track	0.32	0.38	0.27	0.000
Advanced math	0.33	0.31	0.34	0.001
- if attending academic track	0.48	0.50	0.47	0.007
Completed first year in high school	0.93	0.92	0.93	0.242
End-of-year math grade	0.00	-0.06	0.06	0.000
Observations	8 885	4 349	4 536	

Note: Column 1 provides the means of student outcomes for all students. Columns 2 and 3 provide the means for boys and girls, respectively. Column 4 provides the *p*-value for the difference in means between boys and girls. The row "Observations" displays the number of students in the different samples in total, not for each of the variables.

+ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$.

Table 2: Descriptive statistics student, parental and school characteristics

	(1)	(2)	(3)	(4)
	All students	Boys	Girls	<i>p</i> -value
Individual characteristics				
Female	0.51	0.00	1.00	-
Birth month	6.37	6.34	6.41	0.330
Immigrant	0.24	0.24	0.24	0.579
Number of siblings	1.82	1.81	1.83	0.272
Birth order	1.84	1.83	1.85	0.280
Parental characteristics				
Parental income	1203	1198	1209	0.459
Married mother	0.68	0.68	0.67	0.265
<i>Mother's education</i>				
Elementary/middle school	0.14	0.14	0.14	0.993
High School	0.36	0.35	0.36	0.870
College	0.39	0.39	0.39	0.989
Advanced college	0.12	0.12	0.12	0.817
<i>Father's education</i>				
Elementary/middle school	0.15	0.15	0.14	0.394
High School	0.44	0.43	0.45	0.141
College	0.25	0.26	0.25	0.860
Advanced college	0.16	0.16	0.15	0.332
School characteristics				
Students in school	77.61	77.09	78.11	0.107
Students in class	18.50	18.53	18.48	0.565
Classes at school	4.24	4.21	4.28	0.038
Observations	8 885	4 349	4 536	

Note: Column 1 provides the means of individual, parental and school characteristics for all students. Columns 2 and 3 provide the means for boys and girls, respectively. Column 4 provides the *p*-value for the difference in means between boys and girls. The row "Observations" displays the number of students in the different samples in total, not for each of the variables.

+ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$.

Table 3: Balance test

	(1) Teacher mindset	(2) Teacher mindset	(3) Teacher mindset	N
Female	-0.001 (0.005)	-0.001 (0.005)	-0.004 (0.007)	8 885
Birth month	0.004 (0.037)	0.003 (0.037)	0.007 (0.039)	8 885
Immigrant	-0.004 (0.005)	-0.002 (0.004)	-0.003 (0.004)	8 720
Number of siblings	-0.008 (0.011)	-0.008 (0.010)	-0.005 (0.009)	8 720
Birth order	0.002 (0.009)	0.005 (0.008)	0.004 (0.007)	8 723
Parental income	0.027* (0.013)	-0.001 (0.007)	-0.002 (0.007)	8 885
Married mother	0.002 (0.005)	0.002 (0.005)	0.003 (0.005)	8 640
Mother's education	0.006 (0.009)	0.007 (0.008)	0.008 (0.009)	8 465
Father's education	0.001 (0.009)	-0.000 (0.008)	-0.004 (0.007)	8 355
<i>p</i> -value on joint significance	0.715	0.988	0.980	
Fixed effects				
School FE	X	X	X	
Characteristics				
Individual and parental		X	X	
Classroom peers			X	

Note: School fixed effect regressions in columns 1 to 3. The number of observations on the different characteristics is specified in the last column. Standard errors clustered on middle school in all specifications. The dependent variable is teacher mindset, constructed using the difference between the leave-out-mean in the class and in the school. The independent variables of interest are female, birth month, immigrant, number of siblings, birth order, parental income (log), married mother, mother's and father's education. Each coefficient is from a separate regression. The last row reports the *p*-values from performing a seemingly unrelated estimation of all regressions in the column. Individual and parental characteristics in columns 2 and 3 include female, birth month, immigrant, number of siblings, birth order, parental income, married mother, as well as mother's and father's education. Classroom peers' characteristics are leave-out-means of classroom peers' individual and parental characteristics and constructed the same way as the teacher mindset variable. The full sample includes 8 885 students. However, for the regressions on immigrant, number of siblings, birth order, married mother and mother's and father's education, between 165 (immigrant and number of siblings) and 530 (education father) observations are missing. For more specifics about missing variables see [Appendix C](#).

+ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$.

Table 4: Effect of teacher mindset on students' mindset

	(1)	(2)	(3)	(4)	(5)	(6)
	General mindset	General mindset	General mindset	Math mindset	Math mindset	Math mindset
Full sample						
Teacher mindset	0.010 (0.011)	0.009 (0.011)	0.010 (0.011)	0.029** (0.010)	0.028** (0.010)	0.027* (0.010)
Boys						
Teacher mindset	0.003 (0.017)	0.004 (0.017)	0.004 (0.017)	0.020 (0.014)	0.021 (0.014)	0.021 (0.014)
Dependent variable mean	-0.036 (1.047)	-0.036 (1.047)	-0.036 (1.047)	0.001 (1.031)	0.001 (1.031)	0.001 (1.031)
Girls						
Teacher mindset	0.021 (0.013)	0.018 (0.013)	0.020 (0.013)	0.039** (0.013)	0.039** (0.013)	0.038** (0.014)
Dependent variable mean	0.034 (0.952)	0.034 (0.952)	0.034 (0.952)	-0.001 (0.970)	-0.001 (0.970)	-0.001 (0.970)
<i>p</i> -value	0.397	0.501	0.466	0.270	0.322	0.342
Fixed effects						
School FE	X	X	X	X	X	X
Characteristics						
Individual and parental		X	X		X	X
Classroom peers			X			X

Note: School fixed effect regressions, with standard errors clustered on middle school in all specifications. The dependent variable is students' general mindset in columns 1 to 3 and students' math mindset in columns 4 to 6. The treatment variable is teacher mindset, constructed using the difference between the leave-out-mean in the class and in the school. Each coefficient is from a separate regression. The *p*-value is for a *t*-test of significant differences between the coefficients for boys and girls. Individual and parental characteristics include female, birth month, immigrant, number of siblings, birth order, parental income, married mother, as well as mother's and father's education. Classroom peers' characteristics are leave-out-means of classroom peers' individual and parental characteristics, constructed the same way as the teacher mindset variable. The full sample include 8 885 students; 4 349 boys and 4 536 girls.
+ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$.

Table 5: Effect of teacher mindset on students' educational attitudes and expectations

	(1)	(2)	(3)	(4)	(5)
	Attribution	Strategy	Anxious	Expectation	Interest
Full sample					
Teacher mindset	-0.033** (0.011)	0.030* (0.012)	-0.045** (0.011)	0.039** (0.010)	0.037** (0.010)
<i>p</i> -value on joint significance			0.000		
Boys					
Teacher mindset	-0.031+ (0.016)	0.043* (0.017)	-0.052** (0.015)	0.047** (0.015)	0.041* (0.018)
<i>p</i> -value on joint significance			0.002		
Dependent variable mean	-0.173 (1.012)	0.030 (1.013)	-0.284 (0.938)	0.130 (0.973)	0.030 (1.018)
Girls					
Teacher mindset	-0.036* (0.016)	0.015 (0.017)	-0.042* (0.018)	0.032+ (0.017)	0.036* (0.014)
<i>p</i> -value on joint significance			0.088		
Dependent variable mean	0.164 (0.960)	-0.028 (0.986)	0.272 (0.982)	-0.125 (1.009)	-0.028 (0.981)
<i>p</i> -value	0.814	0.213	0.697	0.531	0.831

Note: School fixed effect regressions, with standard errors clustered on middle school. The dependent variable is attribution, strategy, anxious, expectation and interest in columns 1 to 5, respectively. The treatment variable is teacher mindset, constructed using the difference between the leave-out-mean in class and school. Each coefficient is from a separate regression. Each model reports a *p*-value from performing a seemingly unrelated estimation of all regressions in the sample. In the last row of the table we report the *p*-value on the test for different coefficients for boys and girls. All specifications include controls for individual and parental characteristics as well as classroom peers' characteristics. Individual and parental characteristics include female, birth month, immigrant, number of siblings, birth order, parental income, married mother, as well as mother's and father's education. Classroom peers' characteristics are leave-out-means of classroom peers' individual and parental characteristics, constructed the same way as the teacher mindset variable. The full sample in columns 1 and 2 includes 7 090 students; 3 463 boys and 3 627 girls, and 8 885 students in columns 3 to 5; 4 349 boys and 4 536 girls. + $p < 0.10$, * $p < 0.05$, ** $p < 0.01$.

Table 6: Effect of teacher mindset on students' challenge-seeking behavior

	(1) Challenging assignment	(2) Challenging worksheet	(3) Easy worksheet
Full sample			
Teacher mindset	0.011+ (0.005)	0.007 (0.013)	-0.012 (0.010)
Dependent variable mean	0.588 (0.492)	0.000 (1.000)	0.000 (1.000)
Boys			
Teacher mindset	0.023** (0.009)	0.004 (0.018)	-0.019 (0.016)
Dependent variable mean	0.625 (0.484)	0.096 (1.031)	-0.055 (0.999)
Girls			
Teacher mindset	-0.000 (0.008)	0.006 (0.017)	-0.007 (0.016)
Dependent variable mean	0.553 (0.497)	-0.092 (0.961)	0.053 (0.998)
<i>p</i> -value	0.051	0.930	0.631

Note: School fixed effect regressions, with standard errors clustered on middle school. The dependent variable is an indicator for choosing the more challenging assignment in column 1 and standardized variables of challenging worksheet and easy worksheet in columns 2 and 3, respectively. The treatment variable is teacher mindset, constructed using the difference between the leave-out-mean in class and school. Each coefficient is from a separate regression. In the last row of the table we report the *p*-value on the test for different coefficients for boys and girls. All specifications include controls for individual and parental characteristics as well as classroom peers' characteristics. Individual and parental characteristics include female, birth month, immigrant, number of siblings, birth order, parental income, married mother, as well as mother's and father's education. Classroom peers' characteristics are leave-out-means of classroom peers' individual and parental characteristics, constructed the same way as the teacher mindset variable. The full sample includes 7 090 students; 3 463 boys and 3 627 girls.

+ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$.

Table 7: Effect of teacher mindset on students' achievement in middle school

	(1)	(2)	(3)
	Math grade	Math grade	Math exam grade
Full sample			
Teacher mindset	0.031** (0.011)	0.010 (0.021)	0.012 (0.018)
Dependent variable mean	0.000 (1.000)	0.003 (0.998)	-0.000 (1.000)
Boys			
Teacher mindset	0.031* (0.015)	0.005 (0.032)	0.001 (0.027)
Dependent variable mean	-0.074 (0.999)	-0.071 (0.995)	-0.044 (0.977)
Girls			
Teacher mindset	0.031+ (0.018)	0.014 (0.033)	0.021 (0.030)
Dependent variable mean	0.071 (0.996)	0.076 (0.996)	0.045 (1.021)
<i>p</i> -value	0.994	0.873	0.649
Sample restrictions			
Those selected for math exam		X	X
Observations full sample	8 885	3 029	3 029

Note: School fixed effect regressions, with standard errors clustered on middle school. The dependent variable is math grade in middle school in columns 1 and 2 and math exam grade in middle school in column 3. Column 1 includes all students in the sample and the sample size is 8 885 students; 4 349 boys and 4 536 girls. In columns 2 and 3 the sample is restricted to only those selected for math exam at the end of middle school. The treatment variable is teacher mindset, constructed using the difference between the leave-out-mean in class and school. Each coefficient is from a separate regression. In the last row of the table we report the *p*-value on the test for different coefficients for boys and girls. All specifications include controls for individual and parental characteristics as well as classroom peers' characteristics. Individual and parental characteristics include female, birth month, immigrant, number of siblings, birth order, parental income, married mother, as well as mother's and father's education. Classroom peers' characteristics are leave-out-means of classroom peers' individual and parental characteristics constructed the same way as the teacher mindset variable.

+ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$.

Table 8: Effect of teacher mindset on students' choices and achievement first year in high school

	(1) Vocational track	(2) Advanced math	(3) Advanced math	(4) Passed	(5) Passed	(6) Math grade	(7) Math grade
Full sample							
Teacher mindset	0.000 (0.005)	0.009* (0.004)	0.015* (0.006)	-0.004 (0.003)	-0.004 (0.003)	-0.008 (0.011)	-0.019 (0.013)
Dependent variable mean	0.324 (0.468)	0.327 (0.469)	0.484 (0.500)	0.925 (0.263)	0.932 (0.251)	-0.000 (1.000)	0.056 (0.970)
Boys							
Teacher mindset	0.008 (0.008)	0.003 (0.006)	0.014 (0.009)	-0.009* (0.004)	-0.008+ (0.005)	0.006 (0.016)	-0.020 (0.019)
Dependent variable mean	0.382 (0.486)	0.310 (0.463)	0.503 (0.500)	0.922 (0.268)	0.926 (0.263)	-0.064 (0.993)	-0.028 (0.967)
Girls							
Teacher mindset	-0.006 (0.006)	0.014* (0.007)	0.016+ (0.010)	0.001 (0.004)	-0.001 (0.004)	-0.017 (0.017)	-0.020 (0.018)
Dependent variable mean	0.268 (0.443)	0.343 (0.475)	0.468 (0.499)	0.929 (0.258)	0.938 (0.241)	0.061 (1.003)	0.123 (0.967)
<i>p</i> -value	0.160	0.245	0.851	0.039	0.227	0.322	0.987
Sample restrictions							
Those attending academic track			X		X		X
Observations	8 885	8 811	5 932	8 885	6 005	8 419	5 649

Note: School fixed effect regressions, with standard errors clustered on middle school. The dependent variable is vocational track in column 1, advanced math in columns 2 and 3, passed math first year in high school in columns 4 and 5 and math grade first year in high school in columns 6 and 7. Columns 1, 2, 4, and 6 include all students in the sample. The sample size in these columns is 8 885 in the full sample; 4 349 boys and 4 536 girls (columns 2 and 6 have some missing observations due to missing data on the choice of advanced math and math grade for some students). In columns 3, 5 and 7 the sample is restricted to only those attending the academic track in high school, and the sample size in this column is 6 005 for the full sample (columns 3 and 7 have some missing observations for the same reason as columns 2 and 6). The treatment variable is teacher mindset, constructed using the difference between the leave-out-mean in class and school. Each coefficient is from a separate regression. In the last row of the table we report the *p*-value on the test for different coefficients for boys and girls. All specifications include controls for individual and parental characteristics as well as classroom peers' characteristics. Individual and parental characteristics include female, birth month, immigrant, number of siblings, birth order, parental income, married mother, as well as mother's and father's education. Classroom peers' characteristics are leave-out-means of classroom peers' individual and parental characteristics constructed the same way as the teacher mindset variable.

+ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$.

Table 9: Effect of teacher mindset on students' math outcomes first year in high school - bounds

	(1)	(2)	(3)	(4)	(5)
	Advanced math	Passed math (lower bound)	Passed math (upper bound)	Math grade (lower bound)	Math grade (upper bound)
Full sample					
Growth mindset teacher (binary)	0.048** (0.014)				
Teacher mindset		-0.000 (0.002)	0.003* (0.001)	-0.031+ (0.018)	0.007 (0.019)
Boys					
Growth mindset teacher (binary)	0.059** (0.021)				
Teacher mindset		0.002 (0.003)	0.005+ (0.002)	-0.032 (0.031)	0.011 (0.033)
Girls					
Growth mindset teacher (binary)	0.038* (0.019)				
Teacher mindset		-0.002 (0.002)	0.001 (0.001)	-0.029 (0.022)	-0.001 (0.025)
Sample restrictions					
Attend academic track	X	X	X	X	X
Participate in advanced math		X	X	X	X
Observations	5 649	2 693	2 624	2 693	2 624

Note: School fixed effect regressions, with standard errors clustered on middle school. The dependent variable is advanced math in column 1, passed math first year in high school in columns 2 and 3, with the lower bound in column 2 and the upper bound in column 3, and math grade first year in high school in columns 4 and 5, with the lower bound in column 4 and the upper bound in column 5. In column 1 the sample is restricted to those attending the academic track and where the math grade at the end of first year in high school is recorded, and the sample size in this column is 5 649 for the full sample. In columns 2 to 5 the sample is restricted even further, to those participating in the advanced math course, and the sample size in these columns is 2 693 for the full sample. When calculating the upper bounds we exclude students choosing the advanced course because of their teacher, that is 69 students in the full sample. In column 1, the treatment variable is a binary variable taking on the value one if having more of a growth mindset teacher, and zero if not (constructed based on the previous specification of teacher mindset, using 0 as the cutoff). In columns 2 to 5 the teacher mindset variable is constructed using the difference between the leave-out-mean in class and school. Each coefficient is from a separate regression. All specifications include controls for individual and parental characteristics as well as classroom peers' characteristics. Individual and parental characteristics include female, birth month, immigrant, number of siblings, birth order, parental income, married mother, mother's and father's education. Classroom peers' characteristics are leave-out-means of classroom peers' individual and parental characteristics constructed the same way as the teacher mindset variable in the previous specifications.
+ p < 0.10, * p < 0.05, ** p < 0.01.

Appendix A Questions about educational attitudes and expectations

Students' attitudes to getting a bad score on a math test are measured by asking them to:

"Imagine that you get a bad grade on a test tomorrow, how likely are you to think that:"

Attribution: "This shows that I am not very good in math."

Strategy: "I can do better next time if I figure out a better way to work."

which they are to rate on a five-point scale ranging from "Unlikely" to "Very likely". All possible answers being: "Unlikely", "Not likely", "Quite likely", "Likely" and "Very likely".

In addition, the students are asked to which degree, on the same six-point scale as used to map students' mindset, they agree with the following statements:

Anxious: "I become insecure and anxious when thinking about high school math."

Interest: "High school math is going to be interesting."

and to rate:

Expectation: "How do you think you will perform in high school math?"

on a six-point scale ranging from "Very bad" to "Very good". With all possible answers being: "Very bad", "Bad", "Somewhat bad", "Somewhat good", "Good" and "Very good".

Appendix B Questions about challenge-seeking behavior

Hypothetical scenario:

"Imagine that, tomorrow, your math teacher hands out two hand-in assignments. You get to choose which one to do. You're grade will not be affected by the assignment. One choice is an easy review: it has math problems you already know how to solve, and you will probably get most of the answers right without having to think very much. It takes 30 minutes to solve. The other choice is a hard challenge: it has math problems you don't know how to solve, and you will probably get most of the problems wrong, but you might learn something new. It also takes 30 minutes. If you had to choose right now, which one would you choose?"

Easy assignment: "The easy math assignment where I would get most problems right."

Challenging assignment: "The hard math assignment where I would possibly learn something new."

"Make your own worksheet" task:

"We are interested in which type of math assignment first year high school students prefer to work on. You are now given the opportunity to make your own worksheet. At the end of the session you will have the opportunity to answer some of these assignments. On the next few pages there are problems from three different areas in mathematics. Choose between two and six assignments from each area. You can choose from assignments that are:

Very challenging - but you might learn a lot

Somewhat challenging - and you might learn a medium amount

Not very challenging - and you probably won't learn very much

Do not try to answer the math problems. Just click the problems you'd like to try later if there's time."

The areas from which the students are given the opportunity to choose assignments are numbers and algebra, measurement and geometry, and statistics and probability.

Appendix C Missing observations for characteristics

We do experience some missing observations for individual as well as parental characteristics. However, we want to make use of the full sample for the students in which we have outcome measures, and therefore we impute the individual and parental characteristics where these are missing. We do this by re-coding the missing data points for the individual and parental characteristics, and generate an indicator equal to one if the observation is initially missing.

The individual characteristics where we have missing observations are presented in [Table C.1](#) below, and include immigrant, number of siblings, birth order, married mother and parental education. For all missing data points, we re-code the missing observation as the mean of that variable in the student's middle school. In the regressions, we use the variable without any missing observations and also include the indicator for missing on the different characteristics.

When creating leave-out-mean variables for peer characteristics we use the variable without any missing for each of the individual and parental characteristics, that is the imputed variables.

Table C.1: Missing observations for baseline individual characteristics

	Observed characteristics	Missing	Percent missing
Immigrant	8 720	165	1.86
Number of siblings	8 720	165	1.86
Birth order	8 723	162	1.82
Married mother	8 640	245	2.76
Education mother	8 465	420	4.73
Education father	8 355	530	5.97
Full sample	8 885		

Appendix D Test of identifying assumption by gender

Table D.1: Balance test - boys sample (n=4 349)

	(1)	(2)	(3)	N
	Teacher mindset	Teacher mindset	Teacher mindset	
Birth month	0.045 (0.049)	0.044 (0.049)	0.043 (0.052)	4 349
Immigrant	-0.008 (0.007)	-0.006 (0.006)	-0.006 (0.006)	4 268
Number of siblings	-0.002 (0.017)	-0.001 (0.015)	0.001 (0.014)	4 268
Birth order	-0.003 (0.014)	-0.002 (0.013)	-0.005 (0.013)	4 269
Parental income	0.034+ (0.020)	0.006 (0.010)	0.005 (0.010)	4 349
Married mother	0.004 (0.007)	0.004 (0.007)	0.002 (0.007)	4 225
Mother's education	0.013 (0.014)	0.012 (0.012)	0.014 (0.012)	4 161
Father's education	-0.003 (0.015)	-0.006 (0.013)	-0.010 (0.012)	4 097
<i>p</i> -value on joint significance	0.657	0.874	0.856	
Fixed effects				
School FE	X	X	X	
Characteristics				
Individual and parental		X	X	
Classroom peers			X	

Note: School fixed effect regressions in columns 1 to 3. The number of observations on the different characteristics are specified in the last column. Standard errors clustered on school in middle school in all specifications. The dependent variable are teacher mindset, constructed using the difference between the leave-out-mean in the class and in the school. The independent variables of interest is birth month, immigrant, number of siblings, birth order, parental income (log), married mother, mother's and father's education in rows 1 to 8, respectively. Coefficients are from separate regressions. Each column reports a *p*-value from performing a seemingly unrelated estimation of all regressions in the column. Individual and parental characteristics include birth month, immigrant, number of siblings, birth order, parental income, married mother, as well as mother's and father's education. Classroom peers' characteristics are leave-out-means of classroom peers' individual and parental characteristics from our main analyses constructed the same way as the teacher mindset variable in the different specifications.

+ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$.

Table D.2: Balance test - girls sample (n=4 536)

	(1)	(2)	(3)	
	Teacher mindset	Teacher mindset	Teacher mindset	N
Birth month	-0.021 (0.054)	-0.018 (0.054)	-0.009 (0.054)	4 536
Immigrant	-0.002 (0.006)	0.001 (0.006)	-0.001 (0.006)	4 452
Number of siblings	-0.011 (0.018)	-0.012 (0.016)	-0.011 (0.015)	4 452
Birth order	0.010 (0.015)	0.013 (0.013)	0.014 (0.012)	4 454
Parental income	0.023 (0.016)	-0.006 (0.009)	-0.006 (0.010)	4 536
Married mother	0.000 (0.007)	0.001 (0.007)	0.003 (0.008)	4 415
Mother's education	0.000 (0.014)	0.003 (0.012)	0.003 (0.012)	4 304
Father's education	0.001 (0.014)	0.001 (0.012)	-0.000 (0.012)	4 258
<i>p</i> -value on joint significance	0.851	0.975	0.976	
Fixed effects				
School FE	X	X	X	
Characteristics				
Individual and parental		X	X	
Classroom peers			X	

Note: School fixed effect regressions in columns 1 to 3. The number of observations on the different characteristics are specified in the last column. Standard errors clustered on school in middle school in all specifications. The dependent variable are teacher mindset, constructed using the difference between the leave-out-mean in the class and in the school. The independent variables of interest is birth month, immigrant, number of siblings, birth order, parental income (log), married mother, mother's and father's education in rows 1 to 8, respectively. Coefficients are from separate regressions. Each column reports a *p*-value from performing a seemingly unrelated estimation of all regressions in the column. Individual and parental characteristics include birth month, immigrant, number of siblings, birth order, parental income, married mother, as well as mother's and father's education. Classroom peers' characteristics are leave-out-means of classroom peers' individual and parental characteristics from our main analyses constructed the same way as the teacher mindset variable in the different specifications.

+ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$.

Appendix E Grade inflation

We investigate the correlation between teachers with more of a growth mindset and grade inflation by first creating an indicator equal to one if the average grades set by the teacher are higher than the average exam grades for the students in the class, restricting the sample to only students selected for exam. If students not selected to take the exam were in classes with students selected to take the exam, we also label their teacher as grade inflated if the average math grade (set by the teacher) of the students selected to take the exam is higher than the average exam grade. However, given that we need some of the students in the class to have participated in the math exam to construct this variable, the sample size when looking at this is going to be lower (5 296 students).

Second, we also create an indicator for having a teacher with a growth mindset. This indicator is constructed the same way as the variable used when calculating the bounds, with a cutoff of zero on the treatment variable, $\Delta\overline{\text{TM}}_{cs}^{-i}$, labeling about half of the teachers as holding a growth mindset and half of them as holding a fixed mindset.

Using the variables explained above, we find a positive correlation of 0.0465. On the other hand, looking at grade restricted teachers, that is teachers giving lower grades than the exam grade, there is also a positive correlation, of 0.0382. Based on these calculations, we assume that the math grades results in middle school are not driven by the teachers with more of a growth mindset being more lenient in setting grades.

Appendix F Attenuation bias

In order to quantify the degree of attenuation bias, we follow the calculation in Appendix B of Chetty et al. (2011), and use the within-class variation of perceived teacher mindset as an estimate of σ^2 . The estimated variance in the "true" teacher mindset variable, $\hat{var}(tm_{cs} - tm_s)$, is:

$$\begin{aligned}\hat{var}(tm_{cs} - tm_s) &= \frac{(NC - 1)^2}{(NC)^2} \left[\hat{var}(TM_{cs}^{-i} - TM_s^{-i}) - \left(\frac{\hat{\sigma}^2}{N - 1} - \frac{\hat{\sigma}^2}{NC - 1} \right) \right] \\ &= \frac{(59.72)^2}{(61.28)^2} \left[1 - \left(\frac{0.013}{16.61} - \frac{0.013}{59.72} \right) \right] \\ &= 0.949\end{aligned}$$

where tm_{cs} and tm_s is the real teacher mindset within the class and school, respectively. We use the sample harmonic means for NC , $NC - 1$ and $N - 1$ since we are dealing with a sample with varying number of students in each class (N), classes within each school (C) and students within each school (NC) across our sample. The bias is then estimated to be:

$$\begin{aligned}&\frac{\frac{NC}{NC-1} \hat{var}(tm_{cs} - tm_s)}{\frac{(NC)^2}{(NC-1)^2} \hat{var}(tm_{cs} - tm_s) + \frac{\sigma^2}{N-1} - \frac{\sigma^2}{NC-1}} \\ &= \frac{\frac{61.28}{59.72} 0.949}{\frac{(61.28)^2}{(59.72)^2} 0.949 + \frac{0.013}{16.61} - \frac{0.013}{59.72}} \\ &= 0.974\end{aligned}$$

This means that our estimated coefficients are attenuated relative to the real coefficient by 2.6 percent.

Online Appendix A Full regression - math mindset

Table OA.1: Effect on math mindset extended

	(1)	(2)	(3)
Teacher mindset	0.029** (0.010)	0.028** (0.010)	0.027* (0.010)
Indicator for assigned group	-0.001 (0.020)	-0.000 (0.019)	-0.001 (0.019)
Female		-0.000 (0.022)	-0.001 (0.022)
Birth month		-0.003 (0.003)	-0.003 (0.003)
Mother's education=2		0.093* (0.039)	0.092* (0.039)
Mother's education=3		0.189** (0.039)	0.190** (0.039)
Mother's education=4		0.197** (0.050)	0.197** (0.050)
Father's education=2		0.071+ (0.040)	0.072+ (0.040)
Father's education=3		0.157** (0.045)	0.156** (0.045)
Father's education=4		0.181** (0.048)	0.180** (0.048)
Immigrant		0.000 (0.030)	0.001 (0.030)
Birth order		-0.005 (0.015)	-0.006 (0.015)
Number of siblings		-0.000 (0.013)	0.000 (0.013)
Married mother		-0.006 (0.023)	-0.006 (0.023)
Parental income		0.000 (0.000)	0.000 (0.000)
Parental income ²		-0.000 (0.000)	-0.000 (0.000)
Parental income ³		0.000 (0.000)	0.000 (0.000)
Mother's education missing		-0.057 (0.090)	-0.056 (0.090)
Father's education missing		-0.016 (0.062)	-0.015 (0.062)

Continued on next page

Table OA.1 – continued from previous page

Married mother missing	-0.119	-0.124	
	(0.109)	(0.109)	
Birth order missing	0.293	0.296	
	(0.937)	(0.956)	
Number of siblings missing	-0.139	-0.141	
	(0.935)	(0.952)	
Immigrant missing	0.000	0.000	
	(.)	(.)	
Peers' female		-0.104	
		(0.128)	
Peers' birth month		-0.008	
		(0.015)	
Peers' mother's education		0.032	
		(0.084)	
Peers' father's education		0.088	
		(0.060)	
Peers' immigrant		-0.079	
		(0.143)	
Peers' birth order		-0.014	
		(0.063)	
Peers' number of siblings		0.019	
		(0.054)	
Peers' married mother		0.009	
		(0.114)	
Peers' parental income		-0.000	
		(0.000)	
Peers' parental income ²		0.000	
		(0.000)	
Peers' parental income ³		-0.000	
		(0.000)	
Constant	0.000	-0.236**	-0.232**
	(0.010)	(0.078)	(0.079)
R^2	0.032	0.044	0.044
Fixed effects	X	X	X

Note: School fixed effect regressions, with standard errors clustered on school in lower secondary school in all specifications. The dependent variable is students' math mindset. The independent variable of interest is teacher mindset, constructed using the difference between the leave-out-mean in the class and in the school. Classroom peers' characteristics are leave-out-means of classroom peers' individual and parental characteristics constructed the same way as the teacher mindset variable in the different specifications. The full sample include 8 885 students.

+ p < 0.10, * p < 0.05, ** p < 0.01.

Paper III



Involving the teachers: effects of a growth mindset intervention with teacher involvement

May Linn Auestad*

Mari Rege*

David S. Yeager*†

April, 2020‡

Abstract

Social and emotional competencies, like holding more of a growth mindset, are becoming increasingly important for human capital development. Consequently, scientists are trying to figure out how these competencies can be developed on a larger scale through testing specific interventions. To date, none of the growth mindset interventions with positive effects include teachers. We address this gap in the literature by having students complete a growth mindset intervention and accompanying it with assistance to teachers so that they can foster a classroom climate that supports a growth mindset. We recruited 205 teachers in 32 schools and randomly assigned them to a control or treatment group. We find that the intervention instilled a growth mindset in general and in math in particular, a subject area of interest. In addition, the findings suggest that the treatment has a positive effect on a type of challenge-seeking behavior that is important for long-term academic success: enrollment in and completion of the more advanced math course in high school. Involving the teachers led to average effect sizes on course-taking that were almost twice as large as past RCTs involving only students (5pp vs. 3pp previously).

*University of Stavanger

†University of Texas at Austin

‡We are grateful to employees in the Akershus and Rogaland County school districts for valuable suggestions regarding the design and implementation as well as help facilitating the RCT. We are also grateful to Leigh Lauritzen and Nettop-UiS for help setting up the computer program and surveys. The authors acknowledge funding from the *Usay: Shaping students learning expectations* project funded by the Norwegian Research Council (grant number: 260407) led by Professor Mari Rege. In addition, we would like to thank Eric Bettinger, Ingeborg F. Solli, Kjell Arne Brekke and Karen Evelyn Hauge, participants of the Quantitative Forum at the University of Stavanger, and participants of the Ph.D. Workshop in Education Economics for helpful comments. Corresponding author's email: maylinna@gmail.com.

1 Introduction

Evidence suggests that social and emotional competencies are gradually becoming more important for the development of human capital (Heckman & Kautz, 2012; OECD Publishing, 2015; Deming, 2017; Edin et al., 2017). Several studies highlight the value of these competencies in reducing anti-social behavior and the need for social and emotional, as well as cognitive skills in order to achieve positive outcomes later in life (Heckman, 2000; Heckman et al., 2006; Kautz et al., 2014; OECD Publishing, 2015; Deming, 2017; Edin et al., 2017). Therefore, economists concerned with human capital development have started exploring how social and emotional competencies can be developed on a larger scale by building on previous literature or working together with scientists from other fields (e.g. Andersen & Nielsen, 2016; Bettinger et al., 2018; Alan et al., 2019; Rege et al., 2019; Yeager et al., 2019).

One such competence is believing that intelligence and talents can be developed and changed through effort, perseverance, hard work and by using different strategies, or what psychologists refer to as holding a growth mindset (Mueller & Dweck, 1998; Dweck, 2006; Yeager & Dweck, 2012). Well-structured randomized experiments have demonstrated that mindset interventions, teaching students about the brain's malleability, can help develop a growth mindset and improve motivation as well as school outcomes (e.g. Aronson et al., 2002; Good et al., 2003; Blackwell et al., 2007; Yeager et al., 2014; Paunesku et al., 2015; Yeager, Romero, et al., 2016; Bettinger et al., 2018; Yeager et al., 2019). To date, the mindset interventions with positive effects do not include teachers, which could be an important component for scaling the interventions. In addition, involving teachers may enforce the positive effects of mindset interventions by teachers changing their practice or by providing an "affordance" to the interventions (Mueller & Dweck, 1998; Kamins & Dweck, 1999; Kane et al., 2011; Walton & Yeager, in press).

In our study, we address this gap in the literature by testing an intervention with specific teacher involvement. Our hypothesis is that the intervention increases students' beliefs in their own abilities to learn and makes them better at utilizing the learning opportunities that already exist. We expect that in turn, this will make students more likely to choose challenging assignments, show more interest in school activities, have higher expectations in relation to their own performance and increase their achievement and perseverance in school (Mueller & Dweck, 1998; Blackwell et al., 2007; Yeager & Walton, 2011; Burnette et al., 2013; Lavecchia et al., 2016; Yeager, Romero, et al., 2016; Andersen & Nielsen, 2016; Bettinger et al., 2018).

We test this hypothesis in Norwegian high schools using a randomized controlled trial (RCT). First, we invite high school teachers to sign up for the research project before randomly assigning them (and their students) to a control or treatment group. After randomization, teachers in the treatment group are invited to a full day seminar, where we provide them with knowledge about growth mindset as well as tools and guidance to help foster such a mindset among their students. The tools include a web-based program, "U-say", developed in Bettinger et al. (2018) based on the intervention in Yeager, Romero, et

al. (2016). In addition, we include boosters to remind students about growth mindset and basic guidelines for classroom communication practices that help foster the development of such a mindset (Haimovitz & Dweck, 2017). We test whether the intervention led to an increased incidence of growth mindset among students. In addition, we look at the effect on students' challenge-seeking behavior, achievement, perseverance and attitudes by comparing outcomes of participating students in the control and treatment group.

The results show that treated students display more of a growth mindset in general and when it comes to math in particular, compared to students in the control group. We also find evidence suggesting that the treatment has a positive effect on students' challenge-seeking behavior, with significant effects on choosing the more challenging assignment if given a choice. Among students attending the academic track, we also find that treated students are more likely to choose the more advanced math course in high school compared to students in the control group, a choice highlighted as a critical gateway for continuing within the STEM fields later on (e.g. Sells, 1980; Freeman & Aspray, 1999; Ma & Johnson, 2008).

Our paper contributes to the literature in two important ways. First, we contribute to the growing literature focusing on the importance of social and emotional competencies for human capital development (e.g. Heckman & Kautz, 2012; Kautz et al., 2014; OECD Publishing, 2015; Deming, 2017; Edin et al., 2017). We do this by building on Bettinger et al. (2018), and test a growth mindset intervention with teacher involvement. Second, by including specific training for the teachers, we also contribute to the literature identifying best practices in teaching and the impact of teachers on students (Hanushek, 1971; Murnane, 1975; Rockoff, 2004; Kane et al., 2011; Grossman et al., 2013; Chetty et al., 2014a,b; Kraft, 2019). What makes this study unique is that we are using a validated student intervention for both the students and the teachers (Bettinger et al., 2018; Yeager, Romero, et al., 2016), in addition to specific teacher training. However, this means that based on the current study we cannot conclude whether all kinds of teacher training would yield positive effects, for instance if the teachers would support the growth mindset if they made the material themselves.

Moreover, directing attention towards the connection between mindset and theories of motivation as well as developing training material for teachers with respect to their role in fostering a growth mindset among their students is highlighted as important in future research (Rattan et al., 2015; Haimovitz & Dweck, 2017). In addition, our post-intervention measures are conducted several weeks after the intervention is completed, allowing us to better understand how this might affect students' mindset in the longer run than previous studies have done (Yeager & Walton, 2011; Burnette et al., 2013; Bettinger et al., 2018). Finally, we replicate the findings on course-taking from Yeager et al. (2019). Thus, we provide the second, large-sample replication with regards to this outcome.¹

¹Not pre-registered in their study, and thus needed further replication.

The rest of the paper is organized as follows: in [Section 2](#) we present our conceptual framework along with the institutional context. A description of the specific intervention used in this study follows in [Section 3](#), before the experimental design and empirical strategy are presented in [Section 4](#). Assessment and data, as well as procedures follow in [Section 5](#) and [Section 6](#). Results are presented in [Section 7](#), with discussions in [Section 8](#), before finally, we conclude in [Section 9](#).

2 Background

2.1 Conceptual framework

Scientists, primarily within the psychology literature, have for decades been concerned with students' beliefs in their own abilities to learn, and how this might affect their behavior, choices and outcomes at school (e.g. Mueller & Dweck, 1998; Blackwell et al., 2007; Burnette et al., 2013; Yeager, Walton, et al., 2016; Bettinger et al., 2018). These beliefs are commonly referred to as mindset (or implicit theory), and we often talk about two different kinds of mindset - fixed and growth - with a spectrum in between. Several studies show that students leaning more towards the "fixed mindset" end of the spectrum believe that there is nothing they can do to change their intelligence or talents, they think these are fixed traits. Students closer to the "growth mindset" end of the spectrum, on the other hand, believe that intelligence and talents can be developed and changed through effort, perseverance and by using different strategies.

Some studies have also found that students with a higher degree of growth mindset differ from students holding more of a fixed mindset in three important ways. First, they work towards different goals. While the objective of students with a more growth mindset is to learn and to improve their intelligence and talents, students with a more fixed mindset want to have their intelligence and talents validated (Dweck & Leggett, 1988; Haimovitz et al., 2011). Second, they view effort differently. Students holding a higher degree of growth mindset look at effort as something that is necessary in order to improve, while students holding a more fixed mindset see effort as something that undermines their natural abilities (Blackwell et al., 2007). Third, they have different approaches when dealing with mistakes. Students holding a more growth mindset embrace mistakes as a learning opportunity, a chance to learn something new, whilst students with a more fixed mindset view mistakes as a confirmation that they don't have what it takes to succeed, or they might try to cover them up (Blackwell et al., 2007; Haimovitz et al., 2011).

Recent literature demonstrates that it is possible to change students' beliefs in their own abilities to learn, that is move them towards a more growth mindset, using interventions designed for this specific purpose (e.g. Aronson et al., 2002; Good et al., 2003; Blackwell et al., 2007; Paunesku et al., 2015; Yeager, Romero, et al., 2016; Andersen & Nielsen, 2016; Bettinger et al., 2018). The interventions that are successful in doing so teach students about connections in the brain and how these change when working on challenging assignments. They also relate the content to concepts that the students are

already familiar with, like how muscles get stronger with exercise and how computers are connected through networks. Perhaps even more importantly, these interventions emphasize struggles, confusion, being frustrated and making mistakes as a normal part of the learning process. However, what the interventions do not include is the teachers, which could be a critical component for scaling the interventions. Being able to scale effective interventions, is, from an economic point of view, important for human capital development (Heckman, 2000). Moreover, involving teachers may enforce the positive effects of mindset interventions by changing teachers' practices (Mueller & Dweck, 1998; Kamins & Dweck, 1999; Kane et al., 2011; Grossman et al., 2013). Finally, teachers could also provide an "affordance" that allows students to see the growth mindset as legitimate and actionable (Walton & Yeager, in press).

Within the literature focusing primarily on how mindsets are developed, some studies look at the role that daily communication in terms of praise and feedback play with regards to students' mindsets (e.g. Mueller & Dweck, 1998; Kamins & Dweck, 1999; Rattan et al., 2015; Haimovitz & Dweck, 2017). In the classic paper by Mueller & Dweck (1998), the authors show how praising children for their intelligence made them more prone to adapting a more fixed mindset, whilst children who were praised for the process of getting there had a greater tendency to adapt a more growth mindset. In turn, those praised for the process had a greater tendency to choose more challenging assignments, they persisted longer, and performed better than those who were praised for their intelligence. This is also supported in later studies (Gunderson et al., 2013; Pomerantz & Kempner, 2013), and in Rattan et al. (2015) the authors highlight the need for developing training materials for teachers with respect to their role in fostering a growth mindset in students.

In our study, we combine these important stands of scholarship and investigate a growth mindset intervention with specific teacher involvement. We do this by building on existing work by Bettinger et al. (2018), and include teachers in the design of an intervention focused on increasing students' beliefs in their own abilities to learn. We expect that in turn, this will make both teachers and students more likely to adapt more of a growth mindset, and consequently become better at utilizing the learning opportunities that already exist. Subsequently, we expect this will be displayed in students being more likely to choose challenging assignments, show more interest in school activities, have higher expectations about own performance and increase their achievement and perseverance in school (Mueller & Dweck, 1998; Blackwell et al., 2007; Yeager & Walton, 2011; Burnette et al., 2013; Lavecchia et al., 2016; Yeager, Romero, et al., 2016; Andersen & Nielsen, 2016; Bettinger et al., 2018).

2.2 Institutional context

Our sample consists of a selected group of first year students in Norwegian upper secondary schools. Thus, we provide a brief description of important aspects about this level of education along with choices students make during this period that can affect their possibilities later on.

2.2.1 Upper secondary education in Norway

Children in Norway generally start upper secondary education (hereafter referred to as high schools) the calendar year they turn sixteen, after completing ten years of mandatory education, consisting of primary (elementary) and lower secondary (middle) school. The content of mandatory education is determined nationally and is similar for all students.

Prior to starting high school, students apply for one of the many educational programs within the academic or vocational track. Educational programs within the academic track consist of three years of schooling, and are intended to prepare students for starting university or college. In contrast, the main purpose of the educational programs within the vocational track is to prepare students for the labor market, and these consist of three to four years of schooling and training oriented around a specific occupation. In most counties students are free to apply for acceptance into any high school they want to attend within their county of residence. If the number of applicants in a specific educational program in a high school exceeds the number of available slots, acceptance is generally based on the grade point average (GPA) from middle school. Thus, the choice of educational program and high school depends on students' interests and preferences, as well as past performance and where they live (Björklund & Salvanes, 2011). Following the completion of three to four years of high school, students move on to college or join the labor force.

In order to advance to college, students have to go through one of the educational programs in the academic track. This can be done either by attending a program in the academic track for all three years, or by attending one year of intense academic training after finishing a program in the vocational track (Markussen & Gloppen, 2012). In addition, acceptance into some majors in college, especially within the STEM fields, requires the student to have participated in the theoretical math course (hereafter referred to as the more advanced math course) in high school. Therefore, as in many other countries, participation in the advanced math course in high school is also considered a critical gateway for continuing within the STEM fields in Norway (Sells, 1980; Freeman & Aspray, 1999; Ma & Johnson, 2008). This means that the choice of math course in high school represents an important real life outcome when looking at the effect of the intervention.

In addition, drop-out rates in Norwegian high schools are high, and only three in four students finish high school in five years (Statistics Norway, 2019). Students who do not complete their high school education represent an immense cost to society, for instance by having more difficulties in finding a job (Falch et al., 2009). In their report looking at the costs of students dropping out of high school, Falch et al. (2009) also suggest that high school graduates are less likely to use public welfare systems. Thus, reducing high school drop-out rates might lead to decreased income inequality and increased welfare. Therefore, seen from the interest of policymakers, one might wonder if increasing students' beliefs in their own abilities to learn might help reduce high school drop-out rates (Yeager & Walton, 2011; Paunesku et al., 2015; Bettinger et al., 2018).

3 Intervention

The teachers signing up for the project were asked to have six workshops with their students, designed to foster a growth mindset. They received training, a growth mindset curriculum developed for high school students as well as tools and guidance to help them pass the information on to their students and integrate it into their regular classroom practice. Below, we first present the growth mindset curriculum in detail before describing the content and setup of the teacher training.

3.1 Growth mindset curriculum

The growth mindset curriculum, "U-say 2.0", was made up of six separate workshops. The six workshops were spread over 17 weeks, mostly in the fall of 2018, and lasted from 30 to 45 minutes each. In the first and third workshop, both lasting about 45 minutes, the same computer program adapted to the Norwegian high school context and used in Bettinger et al. (2018), building on work by Yeager, Romero, et al. (2016) and Blackwell et al. (2007), was applied. The other four workshops were classroom discussions, lasting about 30 minutes each. See [Figure 1](#) for an illustration of the curriculum.

In the first workshop, ref "Session 1" in [Figure 1](#), the students accessed the web-based computer program, "U-say", through the school district's website. In the program they got to learn about what happens in the brain as you learn new things, why it is important to challenge yourself - especially during high school - and how a growth mindset has helped other students. In addition to looking at pictures, and reading and listening to text,² we asked the students several questions in all parts of the program. The answers were handed in to the teacher at the end of the workshop, and served as a base for the classroom discussion the following week.

Before the second workshop, the first of four classroom discussions, ref "Discussion 1" in [Figure 1](#), the teacher read through the students' answers from the first workshop ("Session 1") and created three to four questions for the students to discuss in the classroom. The teachers were instructed to base the questions on the following themes: different learning strategies, struggling, being frustrated and making mistakes is normal when learning, and motivation to continue even though things are difficult. First, students discussed the questions in small groups, before participating in a full class discussion. The goal of this and the following discussions was to make the students comfortable in reflecting and talking about a growth mindset. In addition, we wanted them to be exposed to other students' thoughts and ways of thinking, and to understand that struggling, being frustrated and making mistakes are a normal part of the learning process.

²All of the text in the program was recorded in order to prevent students losing interest due to having to read the text.

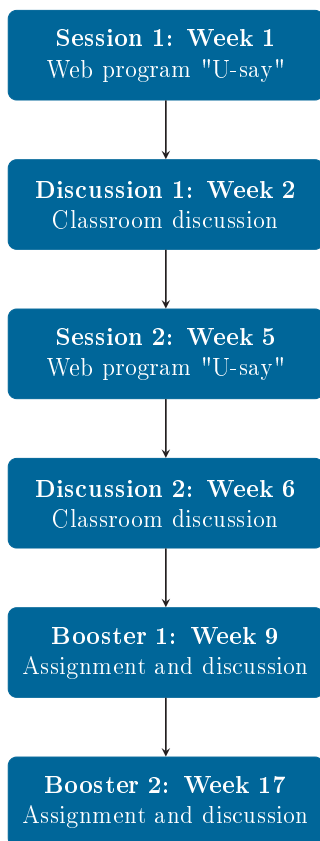


Figure 1: Growth mindset curriculum

The third workshop, ref "Session 2" in Figure 1, consisted of the second session of the web-based computer program, "U-say". Similar to the first session, the students logged on through the school district's website. This time, the students learned about what scientists say about holding a growth mindset, what a growth mindset has helped other people do, examples of famous people who possess a growth mindset and other high school students' thoughts on the growth mindset. As in the first session, students were asked questions throughout the program, which they answered and handed in to the teacher at the end of the workshop.

As in the second workshop, the teachers also read through the students' answers from the web-based program ahead of the second classroom discussion in the fourth workshop, ref "Discussion 2" in Figure 1. Based on the answers, they prepared three to four questions for the classroom discussion. The themes we asked them to base their questions on this

time were what students find important and how to encourage yourself and others to use a growth mindset. The layout and goals were the same as in the previous classroom discussion.

Prior to the final two workshops, ref "Booster 1" and "Booster 2" in [Figure 1](#), we asked the teachers to give the students a hand-in assignment - a booster. After receiving students' answers, the teachers were to read them through and use them as a base for the final two classroom discussions. The layout and the goal of the final two workshops were the same as the previous classroom discussions. The theme of "Booster 1" and "Booster 2" were "cooperation" and "after a holiday" in relation to growth mindset, respectively. In addition to these overall themes, we also asked the teachers to formulate questions based on the following sub-themes: struggling, being frustrated and making mistakes is normal when learning, and the importance of effort, good strategies and knowing when to ask for help.

3.2 Teacher training

Findings from previous literature, looking at how mindsets are naturally developed, suggest that simply handing the teachers a curriculum is not enough, but that instructional practices matter and that it is important to activate teachers' mindset as well as their theories on how to motivate children (e.g. Mueller & Dweck, 1998; Kane et al., 2011; Rattan et al., 2015; Haimovitz & Dweck, 2017). Therefore, we also provided the teachers with training as a part of the intervention.

The teacher training consisted of a full day seminar prior to the implementation of the growth mindset curriculum. During the seminar, the teachers received information about what previous research has found and why we should care about this in the Norwegian context. We also provided them with a direct link between the intervention and the national curriculum, and they learned about the difference between a growth and a fixed mindset. They were also given specific information about the research design.

In addition, during the full day seminar, we went through all six workshops in the curriculum, with specific information about when each workshop was to be implemented. We went through the two web-based sessions, "Session 1" and "Session 2" in [Figure 1](#), in full, and had the teachers answer as if they were doing the program themselves.

Afterwards, we provided them with information about practices in the classroom that substantiate the development of a growth mindset. Particularly, what to keep in mind in order to create a classroom climate that supports a growth mindset, feedback that bolsters the development of a growth mindset as well as the role of learning strategies in promoting a growth mindset. We focused on creating an environment where it is normal to be frustrated and make mistakes when learning, as well as the importance of high expectations, adjusted education, collaboration, displaying and modeling different learning strategies and encouraging students to ask questions and be curious. When giving students feedback, we highlighted the importance of focusing on the process and provide constructive feedback to help students move forward.

At the end of the full day seminar, the teachers practiced creating questions for the classroom discussions using real student answers. Prior to the seminar, the student answers were collected by one member of the research team visiting and performing the first session of the web-based program at a high school in one of the participating school districts. Afterwards, the teachers also answered one of the boosters themselves, and used their answers as the base for practicing one of the booster discussions. The goal of the full day seminar was to make the teachers confident in conducting the growth mindset curriculum with their students, and throughout the day we encouraged them to ask if they had any questions or concerns.

In addition to the information in the full day seminar, the teachers were also given detailed descriptions of all six workshops in the growth mindset curriculum. The description specified when each workshop was to be implemented, how long it would take, the goals of that particular workshop and what learning aids they needed in order to conduct the workshop. In addition, seating arrangements of the students, and what the teachers needed to do prior to and following the workshop were also included in the description. Finally, the week before each workshop, the teachers received an email reminding them about the workshop with the detailed description attached.

4 Experimental design and empirical strategy

Figure 2 illustrates the experimental design. 180 participating classes were randomly split between the control and treatment group using a clustered block design (see Section 6 for specifics). After randomization, the treatment group consisted of 91 classes, and 89 classes made up the control group. Thereafter, teachers of classes assigned to the treatment group participated in the teacher training, consisting of one full day seminar, before they implemented the growth mindset curriculum in their classes the following months (see Section 3 for specifics). The teacher training took place in September 2018, (referred to as "Week 0" in the figure), before the growth mindset curriculum was implemented from the last week of September 2018 through January 2019 ("Weeks 1 - 17"). The participating teachers assigned to the control group continued with business as usual, but received the teacher training and growth mindset curriculum at the start of the following school year ("Next year").

In order to investigate the effect of the intervention, we needed comparable measures for students in the control and treatment group. These were collected by conducting surveys both before and after the intervention, as well as collecting data from government administrative records provided to us by the participating school districts. The pre-intervention survey was conducted before the participating teachers were made aware of their treatment status, the week before the teacher training ("Week -1"). The post-intervention survey was conducted at the end of February 2019, a while after the last workshop in the growth mindset curriculum ("Week 23").

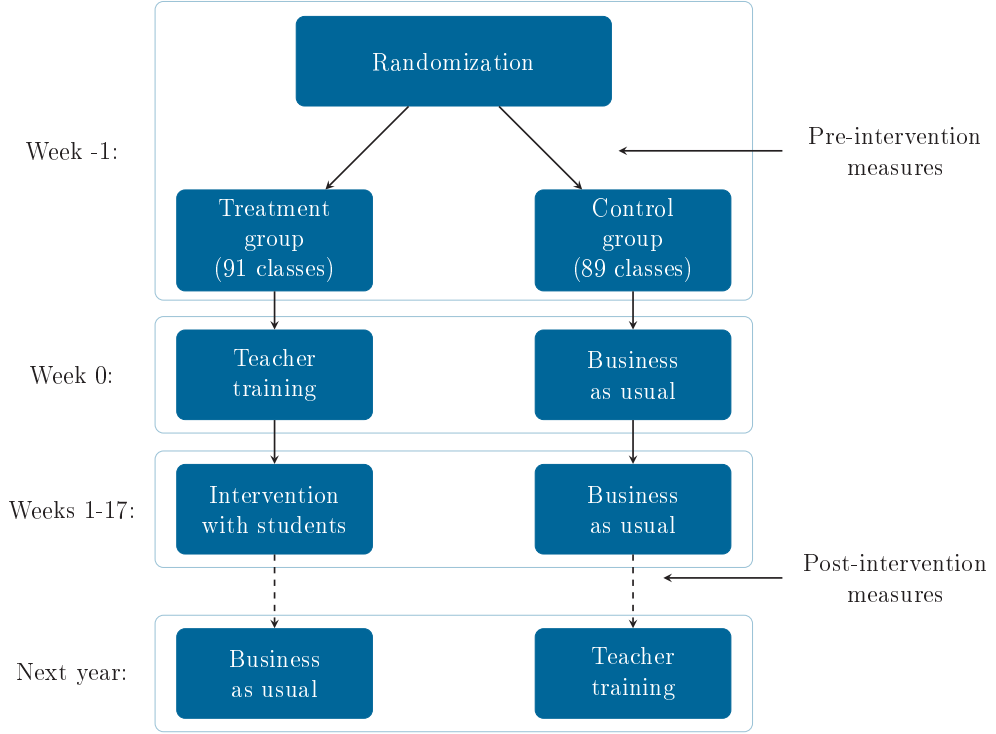


Figure 2: Experimental design

As mentioned, we used a clustered block design when performing the randomization, and consequently we need to adjust for this in our empirical strategy when estimating the effect sizes. Each block consisted of all classes within a separate educational program and school, and the treatment happened at class level. This means that we have to include block fixed effects, that is fixed effects for educational program and school, as well as standard errors clustered at class level in our analyses. This leaves us with the following fixed effect model:

$$(1) \quad y_{ices} = \beta T_{ces} + \zeta X_{ices} + \gamma_{es} + \epsilon_{ices}$$

where y_{ices} is the outcome for student i in class c in block es , where each block represents a separate educational program (e) and school (s). The outcomes of interest in this study are mindset, challenge-seeking behavior, achievement, perseverance, attitudes

and expectations.³ T_{ces} represents an indicator for being treated, equal to one for students in classes receiving the treatment and zero for students in control classes. The coefficient of interest is β , which represents the causal effect of receiving the treatment. X_{ices} is a vector representing controls for students' individual characteristics, whereas γ_{es} denotes block fixed effects, that is fixed effects for students attending the same educational program and school. ϵ_{ices} is an error term specific for each student, capturing unexplained variation in the outcome variable. Robust standard errors clustered at class level are included in all specifications. The set of individual characteristics we include is made up of baseline mindset, GPA and math grade in middle school, as well as indicators for being female, being older than 16 and being an immigrant.

In addition to looking at the effect in the full sample, that is both students attending the vocational track as well as students attending the academic track, we also investigate the effects in subsamples of students based on track in high school as well as GPA from middle school.⁴ We do this because of evidence from previous research that students attending the vocational track as well as students with a lower GPA might respond differently to the treatment than students attending the academic track or students with a higher GPA (Good et al., 2003; Paunesku et al., 2015; Bettinger et al., 2018).

5 Assessments and data

Our sample consists of a selected group of first year high school students attending a vocational track as well as students attending an academic track in two school districts in Norway (for specifics about how the students were selected see [Section 6](#)). For the students who consented to us collecting information about them, we collected data from government administrative records, in addition to conducting student surveys at two points in time; the pre-intervention survey (baseline measures) were conducted in September 2018, while the post-intervention measures were collected in February 2019. In both the pre- and post-intervention surveys we used questions previously used and validated in the mindset literature (Burnette et al., 2013; Yeager, Romero, et al., 2016; Bettinger et al., 2018), described in more detail below.

2 878 students logged on and participated in the "Consent" survey, while 2 623 students gave their consent to participate in the research project and have information about them collected by us. For these students, the school district provided us with registry data. Of the 2 623 students that consented, 74 were missing a class or school identifier, which is crucial for our identification strategy, making us drop them from the sample. This gave us a total of 2 549 students. Of these, 2 364 students participated in the pre-intervention survey, "Survey 1", conducted at the start of the experiment,

³All outcomes, except for perseverance, are specified in the study registration at the American Economic Association (Auestad et al., 2019). Outcomes described in detail in [Section 5](#).

⁴Subsample analyses not included in the study registration, ref Auestad et al. (2019).

before teachers learned of their treatment status. In the spring of 2019, after the students in the treatment group were finished with the intervention, all students in the research project logged on through the school district's website again for the post-intervention survey, "Survey 2". Of the 2 549 students making up the "Consent" sample, 1 768 students participated in this second round of questions. Of these, 1 583 students had also participated in the pre-intervention survey ("Survey 1").⁵

As mentioned in the introduction and conceptual framework, our hypothesis is that the intervention increases students' beliefs in their own abilities to learn and makes them better at utilizing the learning opportunities provided to them. Thus, we want to uncover the effect of the intervention on students' outcomes. Hence, what we are interested in is the effect of the treatment on the treated (TOT), and therefore we use the sample from the post-intervention survey, "Survey 2", when investigating the effects. The reason being that these are the students we know to be treated if placed in the treatment group.⁶

5.1 Variable definitions

5.1.1 Outcome variables

Our primary outcome measures when investigating the effect of the intervention include mindset, challenge-seeking behavior and achievement (Auestad et al., 2019), while our secondary outcomes include perseverance, as well as students' attitudes and expectations towards math. Below we describe each of these in turn.

Students' mindset: We constructed a measure of students' general mindset, as well as their mindset when it comes to math by having them report to which degree, on a six-point scale ranging from "Strongly disagree" (1) to "Strongly agree" (6), they agree with the following statements; (1) "You have a certain amount of intelligence, and you really can't do much to change it" (Mindset 1); (2) "Your intelligence is something about you that you can't change very much" (Mindset 2); and (3) "Being a 'math person' or not is something that you really can't change. Some people are good at math and other people aren't" (Math Measure). We invert all answers in order to have high values reflecting a growth mindset. The first two measures represent two different ways of phrasing the same question, thus in order to construct a measure of students' general mindset, we

⁵This meant that 185 of the students participating in the post-intervention survey had not participated in the pre-intervention survey. However, these students were registered in the participating teachers' class at the time of the post-intervention survey, and therefore we collected consent and answers from these students as well.

⁶In addition, since the teachers had to also initiate the second round of questions, we assume that it is the same kind of teachers in the control group not having their students do this survey as the teachers dropping out of the treatment group. Meaning that the students we are losing from the control group in the second survey are the same kind of students we have lost in the treatment group, which our balance test in [Section 7](#) seems to support. Thus leaving us with similar students in the control and treatment groups when using the post-intervention survey sample.

take the mean of these two (standardized) measures and re-standardize. When creating the math mindset, we standardize students' responses to the third statement, the "Math Measure".⁷

Students' challenge-seeking behavior: In the surveys, we measure both students' attitudes and their behavior towards challenge. Their challenge-seeking behavior is measured by having them create their own worksheet, consisting of easy, medium and/or challenging tasks.⁸ Based on the students' answers, we construct a variable counting the number of challenging tasks chosen, which we then standardize with a mean zero and standard deviation of one. Students' attitudes towards challenge are measured by presenting a hypothetical scenario, where the students are asked to choose between an easy and a more challenging assignment, and where we construct an indicator for choosing the challenging assignment. The exact setting and questions we used are specified in [Appendix B](#).

With regards to challenge-seeking, several studies have highlighted the choice of math course in high school as a critical gateway into continuing within the STEM fields (Sells, 1980; Freeman & Aspray, 1999; Ma & Johnson, 2008; Petersen & Hyde, 2014). This also applies in Norway, with most STEM majors requiring the more advanced math course from high school, making this an important real life choice for students. From the registry data, we therefore include an indicator for choosing the more advanced math course in high school as an outcome when looking at challenge-seeking behavior.⁹

Students' achievement: We are also concerned with students' real life outcomes, and because of this, we included measures for short-term achievement in our design. In particular, we had the students do a math quiz towards the end of the post-intervention survey. Our outcome variable for achievement on the math quiz is standardized number of correct answers. More details about the math quiz are specified in [Appendix C](#). In addition, we also collected information about the students' math grades and GPA at the end of the first year of high school. Thereafter, we created an indicator for passing math, equal to one for all students who attained at least grade 2 in math or who had passed the first year of high school as well as an indicator for completing the first year of high school.¹⁰ Our outcome with regards to GPA was standardized in order to make interpretation easier.

*Students' perseverance:*¹¹ We measured students' perseverance by recording the time spent on the math quiz mentioned above (see [Appendix C](#) for specifics about this exercise). In particular, we recorded how long students used on the first question, the first five

⁷In [Appendix A](#) we present the correlations for the baseline measures, which lay the foundation for grouping the mindset measures this way.

⁸From which they had to solve some randomly selected tasks afterwards.

⁹However, since this is only an option in the academic track, we only include students attending the academic track when looking at this specific outcome.

¹⁰The GPA (and completed), however, includes the grade set by the treated teacher, so this might be influenced by potential differences in grading schemes as a function of the intervention. Therefore we only report these measures for the sake of completeness.

¹¹Outcome not included in the the study registration, ref Auestad et al. (2019).

questions and all ten questions. When investigating the effects with regards to time spent, we use logarithms.

Students' attitudes and expectations: Students' attitudes and expectations were measured by asking; (1) if the student would attribute a bad test score to not being good in math ("Attribution"); (2) if the student believes that using another strategy could have increased the score ("Strategy"); (3) if becoming anxious and insecure when thinking about math ("Anxious"); (4) if the student believes high school math will be interesting ("Interest"); and (5) the student's expectation about own performance in math in high school ("Expectations"). Answers to each of the survey questions are standardized with a mean of zero and a standard deviation of one. The exact questions used are specified in [Appendix D](#).

5.1.2 Control variables

Control variables include standardized variables for baseline mindset, GPA and math grade at the end of middle school, as well as indicators for gender (female), being older than 16 and being an immigrant. We also include indicators for missing observations on control variables (see [Appendix E](#) for specifics about how we deal with these). The baseline mindset is constructed the same way as the outcome variable on mindset, using baseline measures from the pre-intervention survey. The rest of the control variables are from government administrative records provided to us by the participating school districts.

6 Procedures

In cooperation with two counties in Norway, referred to as the participating school districts, we started by inviting all homeroom teachers responsible for first year high school students to participate in the research project. We sent informational emails before the summer break, with the registration deadline set to the beginning of the following school year (2018/2019). We sent reminders at the start of the school year. As mentioned in the institutional context in [Section 2](#), Norwegian high schools are divided into tracks, and each track consists of several educational programs. Students sort into these programs and schools based on interest, preferences, place of residence and grades from middle school, and in order to have comparable groups of students we encouraged teachers to sign up in pairs. This meant that if there was interest in the project, we encouraged at least two teachers (responsible for separate classes) within each educational program and school to sign up. Participation for the teachers was voluntary, but they committed to performing both the surveys and the required tasks with their students when signing up. However, we did not have any sanctions if the teachers decided not to participate. 205 teachers responsible for 180 different classes in 32 schools signed up. Some teachers later withdrew and we discuss possible concerns regarding this later when looking at attrition.

The classes of participating teachers were grouped into blocks, where each block represented classes within the same educational program and school. After doing this we had a total of 89 blocks. The blocks were used as a means of randomization, meaning that within each block, we randomized classes to control and treatment group, also referred to as a clustered block design. By doing this, we had at least one class in the treatment group and one class in the control group in all blocks consisting of at least two classes. If, for some reason, there was more than one homeroom teacher in a class, these teachers were seen as one when performing the randomization. That is, randomization was done at the class level, and both teachers were placed in the same condition. We randomized by assigning each class with a random number between 0 and 100, using a computerized random number generator. Some schools also had more than a pair of participating classes within an educational program, and in those cases we also considered the gender of the teachers when performing the randomization, as this has been shown to influence student outcomes (Dee, 2007). This means that if there were two female teachers and one male teacher within a block, we randomized between the two female teachers.

We sorted from highest to lowest based on the randomly generated number, and assigned the first teacher in a block to the treatment group, the second to the control group, the third to the treatment group, the fourth to the control group, and so on. In the event of an uneven number of classes within a block, that is classes without any pairings (the first,¹² the third, the fifth, etc.), these classes were placed in a residual group. After the first round of randomization within blocks, classes placed in the residual group were randomized as though they were in the same block, also here taking gender of the teacher into consideration. That is, the first was placed in the treatment group, the second in the control group, the third in the treatment group, the fourth in the control group, etc. This was done in order to get approximately the same number of classes in the control and treatment group. The randomization was carried out by a computer in the office, with a witness present. After randomization, the treatment and control group consisted of 103 teachers responsible for 91 classes, and 102 teachers responsible for 89 classes, respectively.

Participation in the field experiment was voluntary for the teachers, however for the students of participating teachers, participation was mandatory as a part of the school instruction, but the students had to give their consent to us collecting information about them. Before the teachers were made aware of their treatment status, we collected consents from the students in the participating classes, and also had them perform the pre-intervention survey. To access all the surveys in the project, the students used their own computers, and logged on using their unique username and password administered by the school district, before they were redirected to the correct survey. During the redirection, students' usernames were replaced with a unique number for each student, making

¹²Despite the fact that we encouraged teachers to sign up in pairs, we did not deny teachers without pairings to sign up, and therefore there are going to be some single class blocks in our sample. 29 of the initial 180 classes were in single class blocks.

it impossible to identify any single student in the data without an identification key. The identification key was safely stored with a trusted employee at the respective school district, never to be seen by anyone in the research team. Later, this unique student number was matched to the administrative records using the identification key, before the de-identified data were provided to us.

The first time the students logged on through the school district's website, in September 2018, they received information about the research project, and had to give their consent to us collecting information about them. For most students (93 percent of those who consented), this was done at the start of the pre-intervention survey (or "Survey 1"). However, due to changes in classes after the initial survey, we also allowed students to consent and take the post-intervention survey if they had not participated in the first round of questions. That is, all students participating in either the pre- or post-intervention survey had to go through the consent survey. Of the 7 percent giving their consent before the post-intervention survey, 44 percent were in treated classes. In total, we received consent from 2 623 students or 91 percent of the students logging on to the consent survey.

In the pre-intervention survey, we asked the students questions to map their mindset, as well as their attitudes and expectations towards math. As the students finished the pre-intervention survey, their teachers were informed about their treatment status. Afterwards, together with the participating school district, we invited all homeroom teachers responsible for a class assigned to the treatment group to a full day seminar, one in each of the two districts. At the start of the seminar, we collected signed confidentiality agreements from the teachers stating that they were not to distribute any of the material they had access to through the intervention in the current school year. In addition, during the seminar, the teachers received all the information they needed in order to implement the growth mindset curriculum with their students, which they did over the following four months.

After the classes assigned to the treatment group were finished with the growth mindset curriculum, we had all participating teachers, that is the teachers in both the control group and the treatment group, implement the post-intervention survey (or "Survey 2"). This was done at the end of February 2019, several weeks after the teachers and students in the treatment group were finished with the intervention. In the post-intervention survey, students were asked the same questions as in the pre-intervention survey, in addition to some new questions intended to detect challenge-seeking behavior and achievement. In addition, after the school year had ended, the participating school districts provided us with data from administrative records for the students who consented. From these registry data we use the following variables in our analyses: unique number for school, class and educational program in high school, GPA and math grade from middle and high school, high school track (indicator for vocational), gender (indicator for female), immigrant and age (indicator for being older than 16).

All sessions, both the surveys and the growth mindset curriculum, were implemented in the students' classroom during regular school hours. Similar to when doing the survey, the students also used their own computer, username and password when accessing the

web-based program in workshop one and three of the growth mindset curricula. For these two workshops, the students were also instructed to bring their headsets, as the text in the program is also recorded. Only students registered in a class assigned to the treatment group were able to access the web-based program. For the four classroom discussions, the students did not need any web-based resources, computers or headsets.

The teachers assigned to the control group continued with business as usual after having their students do the consent and pre-intervention survey until the post-intervention survey. That is, they only implemented the consent, pre- and post-intervention surveys with the students in their class. The following school year (2019/2020), they received the same information and material as the teachers in the treatment group had received the year before. The students of the teachers in the control group, however, might not benefit from the control group teachers getting the treatment, since teachers do not always follow students through high school.

Despite the fact that there are a lot of benefits to working with real life situations, we are dealing with people, and their behavior is potentially going to affect the experiment (Creswell & Creswell, 2017; Yeager & Walton, 2011; Cartwright, 2012). In our study, we do experience a high attrition rate, with about 30 percent of the students in our "Consent" sample not participating in the post-intervention survey, and therefore not being present in our "Survey 2" sample (for more detailed numbers see [Section 5](#)). This was one of the biggest challenges in our study, and we acknowledge several possible sources for this attrition. First, since we are dependent on teachers both when performing the surveys and the intervention, a lot of the attrition can be attributed to teachers withdrawing and thereby preventing entire classes from participating in the project. Second, because Norwegian high schools have a big problem with absence and drop-out (Bettinger et al., 2018; Statistics Norway, 2019), we do experience some attrition from this as well, with students not being present at school when the surveys are conducted. Third, because we base most of our outcomes on survey questions, we encounter some attrition because students skip some questions in the survey. Since we already have a selected sample, in that teachers select into treatment, drop-out from our sample only causes problems if we lose different kinds of students from the treatment group compared to the control group. If this happens, the groups are no longer going to be balanced at baseline, which is what random assignment assures, and we will no longer be comparing similar students. See [Appendix F](#) for a discussion about the attrition, as well as a detailed attrition analysis.

Treatment compliance was assessed using a questionnaire for the teachers at the same time as the post-intervention survey for the students. In the survey, we asked the teachers if they had completed each of the specific workshops in the growth mindset curriculum with their students, and if not, why this was. We received answers from 68 teachers responsible for 60 different classes (93.75 percent of the treated classes in our post-intervention sample). Of the teachers who reported back, 80 percent had performed all of the workshops with their students and 3 percent of the teachers reported that they had not performed any of the workshops. 90 percent of the teachers reported that they had completed the first four workshops with their students. The reasons for not

performing all of the workshops were due to time constraints (85 percent) and lack of interest from the students (15 percent). In addition, 66 of the teachers also answered some additional questions, where 80 percent reported that they found the intervention meaningful and that their students benefited from participating. Moreover, 60 percent reported that their students found it enjoyable, 90 percent reported that they would continue teaching their students about a growth mindset and 75 percent reported that they wanted to continue using the growth mindset curriculum in the years to come.

7 Results

7.1 Descriptive statistics and balance test

The high attrition rate in our study (ref [Section 6](#), as well as [Appendix F](#)), could cause problems with balance in the sample we base our analyses on. Therefore, in addition to the attrition analyses done in [Appendix F](#), we test for balance by regressing students' individual characteristics against the treatment status, using school and educational program fixed effects and robust standard errors clustered at class level. [Table 1](#) reports descriptive statistics for the control and treatment groups in the columns labeled "Control" and "Treat", and the balance test in the column labeled "Diff". Descriptive statistics and the balance test for the "Consent" sample are reported in columns 1 to 3, the "Survey 1" sample in columns 4 to 6 and the "Survey 2" sample in columns 7 to 9 (for specifics about the samples, ref [Section 5](#)). Doing this, we only find one significant coefficient, which might suggest moderate imbalance between the control and treatment group (older than 16 at 10 percent significance level in the "Consent" sample and 5 percent in the "Survey 1" and "Survey 2" samples). However, this variable was already imbalanced from the start, and did not change very much because of the attrition. In addition, the difference here can be attributed to a large number of older students (7 or 8) in a few classes.

Based on the attrition analyses in [Appendix F](#), as well as the balance test in [Table 1](#), we conclude that attrition in the post-intervention survey did not lead to sizable significant differences between the two groups, that is the treated and the control group. Therefore we conclude that the sample is quite well-balanced when focusing on the students participating in the post-intervention survey or the "Survey 2" sample. However, because of some large, but not significant, differences (on for instance the baseline mindset measure), we do present the estimated effects both with and without control variables in the analyses below.

7.2 Treatment effects

7.2.1 Primary outcomes

Our primary outcomes include students' mindset, their challenge-seeking behavior as well as achievement, described in more detail in [Section 5](#). Estimated treatment effects, as well as number of observations and adjusted R^2 are reported in Panel A of [Tables 2 to 4](#) for the full sample. To address the concern about imbalance on some of the characteristics, we include two models for each of the outcomes in all tables; one where we only include the fixed effects and one where we include both the fixed effects and control variables, indicated in the last line of the table.

Our first outcome of interest, students' mindset, includes both their general mindset and their mindset when it comes to math in particular, and the estimated coefficients are presented in [Table 2](#). We find positive and significant effects of the treatment both with regards to students' general mindset, reported in columns 1 and 2, and their mindset when it comes to math, reported in columns 3 and 4. The size of the estimates are 20.2 percent of a standard deviation for the general mindset and 20.5 percent of a standard deviation for the math mindset in our preferred specification, where we include both fixed effects and control variables. Corresponding numbers are 16.7 and 19.8 when only including fixed effects. The difference between the estimated coefficients particularly for mindset, in columns 1 and 2, can be attributed to the small imbalance we find on individual characteristics in the balance test, as noted above (illustrating the importance of including control variables in the regressions).

In [Table 3](#) we report the estimated effects on students' challenge-seeking behavior and attitude. The treatment effect with regards to challenge-seeking behavior is reported in columns 1 and 2 of the table, where we find a positive estimate of 8.2 percent of a standard deviation in our preferred specification (significant at the 10 percent level), when looking at the number of challenging assignments chosen. For students' attitudes towards challenge, reported in columns 3 and 4, we find that treated students are 5.2 percent more likely to choose the more challenging assignment (significant at the 5 percent level). The last outcome we include with regards to challenge-seeking is the choice of the advanced math course in high school, reported in columns 5 and 6. Here, we restrict the sample to only include students attending the academic track, since students attending the vocational track do not have the option of selecting which math course to participate in. In our preferred model, presented in column 6, we find a significant positive treatment effect of 5.7 percent.¹³

¹³One concern, however, is that pushing students towards more challenge-seeking behavior in math makes them more likely not to pass. Therefore, we also looked at the treatment effects on getting a pass grade in math in the first year of high school and did not find any indications that the treatment had any effect on this outcome. In addition, we investigated the effect on an indicator for choosing and passing the advanced math course (that is, an indicator equal to one for students who chose to participate in and passed the advanced math course), and found a positive treatment effect of 5.4 percent in our preferred

Students' achievements are measured using a math quiz consisting of ten math assignments towards the end of the post-intervention survey. In addition, for the sake of completeness, we include GPA as well as an indicator for completing the first year in high school. The estimated treatment effects are reported in [Table 4](#). We do not find any significant treatment effects for any of the outcomes when looking at achievement in the full sample.

7.2.2 Secondary outcomes

Our secondary outcomes include perseverance¹⁴ and students' attitudes and expectations towards math, and estimated coefficients are reported in [Appendix G](#).

In [Appendix Table G1](#) we report the estimated effects on time spent on the first question in columns 1 and 2, the first five questions in columns 3 and 4 and all ten questions in columns 5 and 6. When looking at the estimated coefficients for the overall sample, they are pointing in the opposite direction of what we expected, indicating a negative effect of the treatment on perseverance. However, none of the estimated coefficients are significantly different from zero when looking at the full sample.

Effects on students' attitudes and expectations are reported in [Appendix Table G2](#). Here we only report our preferred specification. The only significant treatment effect for the full sample is on interest, with a positive significant estimate of 9.6 percent of a standard deviation (significant at the 10 percent level). This indicates more interest in math among treated students. Based on the estimated effects, treated students seem less likely to attribute a bad score on a test to their abilities and more likely to believe that using a different strategy would have helped. In addition, they also seem to have higher expectations and be a little more anxious when thinking about high school math (however, none of the estimated effects other than on interest are significantly different from zero).

7.2.3 Subsample analyses

Previous studies have pointed out the fact that students might respond differently to an intervention focusing on increasing students' beliefs in their own abilities to learn (Good et al., 2003; Paunesku et al., 2015; Bettinger et al., 2018), and because of this, we also investigate the effects in subsamples of students.¹⁵ First, we divide the students based on track in high school, that is we look at the effects separately for students attending the vocational and academic tracks. We lose a lot of power in our analyses when doing this, since this basically means cutting the number of blocks, and therefore also our sample, in

model (i.e. very similar to the estimated effect of 5.7 percent presented here). Analyses available upon request.

¹⁴Outcome not included in the study registration, ref Auestad et al. (2019).

¹⁵Subsample analyses not included in the study registration, ref Auestad et al. (2019).

half. Second, within each track, we also divide students based on their pre-intervention GPA, that is we include an indicator for high GPA from middle school in the analyses, which we interact with all other variables in the regression. We consider students with a GPA above the 50th percentile as having a high GPA.¹⁶

Tables 2 to 4 present the subsample analyses for our primary outcomes, while Appendix Tables G1 and G2 present the subsample analyses for our secondary outcomes. Panels B and C in each table report the estimated effect of the treatment on students attending the vocational and academic track, respectively. Panel D reports the estimated treatment effects based on GPA within the vocational track, while panel E reports the estimated effects based on GPA within the academic track. In addition to the estimated effect of the treatment, as well as the interaction between high GPA students and treatment in the panels where we use interactions, we also include number of observations and adjusted R^2 in all columns and panels. The first column under each outcome for our primary outcomes, in Tables 2 to 4, as well as our secondary outcomes on perseverance in Appendix Table G1, reports the estimated coefficients using only fixed effects, while the second column reports the estimated coefficients from our preferred specification using both fixed effects as well as control variables, indicated in the last line of the tables. For our secondary outcomes on students' attitudes and expectations, reported in Appendix Table G2, we only include our preferred specification.

When looking at the effect in our subsample analyses with regards to mindset, in Table 2, we find a positive treatment effect of 10.6 percent of a standard deviation for the general mindset and 14.2 percent of a standard deviation for the math mindset among students attending the vocational track (significant at the 5 percent level). Among students attending the academic track, the treatment effects are larger, with 30.7 and 26.4 percent of a standard deviation, for the general and math mindset, respectively. When also taking pre-intervention GPA into consideration, we do not find any significant effects when looking at the vocational track students. However, for low GPA students attending the academic track we find positive treatment effects of 42.5 and 36.7 percent of a standard deviation for the general and math mindset, respectively. For high GPA students attending the academic track, the corresponding estimates are 22.4 and 19.9 percent of a standard deviation. For general mindset, the estimated coefficients for low and high GPA students within the academic track are also significantly different at the 10 percent level.

With regards to students' challenge-seeking behavior, reported in Table 3, we find indications that the treatment affects the challenge-seeking behavior and attitudes of students attending the vocational track. Here we find positive treatment estimates of 11.6 percent of a standard deviation when looking at number of challenging assignments chosen and 5.5 percent when looking at choosing the more challenging assignment in

¹⁶When creating this variable, students whose GPA from middle school was missing were put in the category of low GPA students. In addition, we also include robustness tests where we exclude them entirely from the sample, in order to investigate if this affects the results in Online Appendix A. It does not.

our preferred model (however, both are only significant at the 10 percent level). We do not find any significant effect of the treatment when looking at students attending the academic track, other than the effect on the choice of math course in high school mentioned when looking at the primary outcomes above. When taking GPA in middle school into consideration, we do not find any significant effect. We do, however, find indications of positive effects on choosing more challenging assignments for low GPA students within the vocational track. The effect on choosing the more advanced math course seems to be coming from both the low and high GPA students attending the academic track.

When looking at achievement in the subsamples, reported in [Table 4](#), we find indications that students attending the vocational track might be negatively affected by the treatment, and that students attending the academic track might be positively affected. None of the estimated coefficients with regards to achievement in the subsamples based on track are significant. When also taking GPA from middle school into account, we find indications that low GPA students within the vocational track seem to be negatively affected by the treatment, and that high GPA students seem to be positively affected with regards to the math quiz and GPA in high school. The estimates are, however, not significant. Among students attending the academic track, we find that low GPA students seem to be negatively affected (significant at the 10 percent level) and that high GPA students seem to be positively affected in terms of completion, with a negative effect of 5.1 percent for low GPA and a positive effect of 2.6 percent for high GPA students (significantly different from each other). For the math quiz, we find indications of positive effect on low GPA students and an effect closer to zero for high GPA students. Lastly, for the GPA in the first year of high school, we find indications of a negative effect for low GPA students and a positive effect for high GPA students.

With regards to perseverance on the math quiz, reported in [Appendix Table G1](#), we find indications of negative treatment effects among vocational track students, which seem to be coming from both low and high GPA students (also significant at the 10 percent level when looking at the low GPA students). With regards to academic track students, however, we find indications of positive treatment effects, which seem to be coming primarily from low GPA students. However, none of the estimated effects with regards to time spent are significantly different from zero for the academic track students.

Subsample analyses with regards to students' attitudes towards math are reported in [Appendix Table G2](#). Here we find indications that students attending vocational and academic tracks are affected by the treatment in the same direction but to a different extent. However, the only significant estimate when looking at the subsamples based on track is a positive effect on interest for students attending the vocational track (significant estimate of 16.7 percent of a standard deviation, significant at the 10 percent level). In addition, there is a significant positive effect on expectation of 16.5 percent of a standard deviation among low GPA students attending the academic track (significant at the 5 percent level).

8 Discussion

In this study we build on previous literature (especially Bettinger et al., 2018) and investigate if it is possible to change students' mindset by including teachers in an intervention focused on increasing the prevalence of a growth mindset among students. We do this by providing knowledge to teachers about the growth mindset and give them tools to help pass the information on to their students. In addition, we have the teachers initiate a post-intervention survey several weeks after the treatment group is finished with the intervention. Thus, we investigate whether changes in mindset persist over a longer time period than previous studies have done (e.g. Yeager & Walton, 2011; Bettinger et al., 2018).

Similar to previous studies, when investigating the effects of providing students with short sessions on the growth mindset without involving teachers, we find that it is possible to change students' beliefs in their own abilities to learn, i.e. their mindset, using a mindset intervention (Blackwell et al., 2007; Yeager, Romero, et al., 2016; Bettinger et al., 2018; Yeager et al., 2019). We also find that the changes persist. In addition, we find indications that this can be transferred to changes in challenge-seeking behavior, in that treated students show more of a positive attitude towards challenge, in line with findings in Mueller & Dweck (1998) and Blackwell et al. (2007). With regards to achievement and perseverance, the estimated effects when looking at the full sample are very small and not significantly different from zero. However, one of the problems we face is that we do not have sufficient power to detect the effect size for these outcomes given the sample we have. We are therefore not able to conclude that the intervention had any affect on these outcomes as previous studies have done (Blackwell et al., 2007; Yeager, Walton, et al., 2016; Bettinger et al., 2018; Yeager et al., 2019).

Students in Norwegian high schools sort into different tracks based on interest, preferences and grades from middle school, meaning that students attending the vocational track might be very different from students attending the academic track. Therefore there have been some suggestions that they might respond in different ways to this kind of intervention. Because of this, we also investigate the effects by dividing the sample based on track in high school when doing the analyses, and find indications in support of this belief.¹⁷ However, we lose a lot of power in our analyses when doing this, since we are basically cutting our sample in half, making it even more difficult to detect significant effects of the treatment.

In addition, previous research suggests that students respond differently based on their initial mindset, which tends to be correlated with GPA (Good et al., 2003; Paunesku et al., 2015; Bettinger et al., 2018). Therefore, we also investigate the effects in subsamples based on GPA from middle school within the different tracks. We do this by including

¹⁷Subsample analyses not included in the study registration, ref Auestad et al. (2019).

an indicator for high GPA students, which we interact with the dependent variable of interest and all control variables.

Our analyses show that the treatment has a positive effect on students' mindset of between 10 and 15 percent of a standard deviation for students attending the vocational track, whereas for students attending the academic track the effect is closer to 30 percent of a standard deviation. In addition, for vocational track students, the effect seems to be similar for low and high GPA students, but for academic track students, the low GPA students show a larger effect than the high GPA students.

For challenge-seeking outcomes, we find that students attending the vocational track seem to be more affected by the treatment, with positive significant estimates for all outcomes (however, only significant at the 10 percent level). For academic track students, we also find positive treatment effects, but the only significant effect is with regards to the choice of the more advanced math course in high school (significant at the 5 percent level). Students attending the vocational track are not able to choose which math course to participate in, hence we do not observe this outcome for vocational track students. Unlike many of the other outcomes, the choice of math course is a real life choice, and many studies have highlighted this as a critical gateway into continuing within the increasingly important STEM fields later on (Sells, 1980; Freeman & Aspray, 1999; Ma & Johnson, 2008). This is also the case in Norway, with most STEM majors requiring the more advanced math course from high school.

One concern, however, when looking at participation in the advanced math course, is that treated teachers may have encouraged students to take the advanced course because of the information given in the teacher training. That is, we do not know for sure that the students themselves made the choice to participate in the advanced course. However, given that we are treating the homeroom teacher (not the math teacher), we have reason to believe that the teachers do not have perfect information about how students perform in math, and therefore which students should participate in the advanced course. In addition, since we had the teachers sign a confidentiality agreement during the teacher training, we assume that they did not share the information given with other teachers, like the students' math teachers. However, of the treated teachers who answered the post-intervention teacher survey, 18 percent reported that they were teaching math to at least some of the students in their "base" class.¹⁸

Moving on to looking at achievement and perseverance, we also uncover some interesting indications when dividing the sample based on track in high school. However, again, not many are significant. We do find, for instance, that treated students attending the

¹⁸In high school, students taking advanced and regular math are usually in the same "base" class, meaning that they have the same homeroom teacher, i.e. the teacher that is treated, but they go to separate classrooms for their math lessons (though with different math teachers). This also means that since we are treating the homeroom teachers, not the math teachers, the students still remain in the treated teacher's class even if they change from the regular to the advanced math course.

academic track do better on the math quiz, and there are also indications that this might be because they persevere with the questions for longer, with positive estimates both on the number of correct answers and on the time spent on the math quiz. In addition, we find opposite effects for students attending the vocational track, with negative estimates on the number of correct answers on the math quiz and time spent.

When taking pre-intervention GPA into consideration, we find indications of negative treatment effects for low GPA students attending the vocational track, positive effects for low GPA students attending the academic track and estimates very close to zero for high GPA students on both tracks when looking at achievement in the math quiz. With regards to time spent, we find indications that treated, low GPA students attending the academic track persevere with the questions for longer. When looking at high school GPA, we also find both positive and negative effects in different samples, however, none of which are significantly different from zero. With regards to completion, we find a negative effect of the treatment on low GPA students and a positive effect on high GPA students attending the academic track. Summing up, we do present evidence that the treatment changed students' mindset and their attitudes towards challenge, as well as indications that it also changed achievement and perseverance among subsamples of students.

However, this is an RCT and we are dealing with people, therefore their behavior could potentially affect the experiment. This is to some extent illustrated by the immense attrition in this study, as mentioned in [Section 6](#) and discussed in more detail in [Appendix F](#). In addition to attrition, we also encounter other challenges in this RCT. The first one being diffusion of treatment (Cartwright, 2012), or spillovers, which is highly likely to happen since we are randomizing within schools (that is, we have both treated and control classes attending the same schools). As with attrition, we face several possible scenarios here, with spillovers between teachers, between teachers and students, and between students. We try to deal with this by having the teachers in the treatment group sign a confidentiality agreement stating that they are not to distribute any of the material they get access to through the intervention. However, because we are dealing with a real life situation, we cannot restrict teachers or students from talking to each other or teachers from using what they learn from the project with other students. That is, if teachers assigned to the treatment group in fact teach students assigned to the control group in some subject, even though they are not their homeroom teacher. It is therefore impossible to remove spillovers completely. Due to the fact that a big part of the intervention is workshops, and that control students won't get this information, the possible spillover concern relates to changes in behavior among teachers and students in the treatment group because of the information given in the teacher training and growth mindset curriculum. Thus, the estimated effects might be biased downwards compared to if we were able to completely remove spillovers from the experiment.

Second, compensatory or resentful demoralization and compensatory rivalry are also possible concerns in our experiment. Compensatory or resentful demoralization happens when participants who are not getting the treatment become discouraged, and as a result perform worse on the outcomes measured (Creswell & Creswell, 2017). This is a known

problem when dealing with RCTs since we need to have both a treatment and a control group. One possible solution could be blinding the participants to their treatment status or providing the control group with some sort of placebo. However, most teachers already feel like they do not have enough time and therefore we are not too concerned with the control group feeling like they are losing out. With regards to blinding the participants, this was never really an option given the extent of the intervention. We try to deal with this by having the pre-intervention measures collected before the participants get to know their treatment status, and by collecting the post-intervention measures when the control group has almost forgotten about the research project.

Compensatory rivalry, on the other hand, happens if the participants in the control group change behavior because they are placed in the control group, i.e. they start working harder, try more, etc (Creswell & Creswell, 2017). Our main concern with regards to this is that teachers in the control group start looking for information about the growth mindset themselves because they really want their students to benefit from this and are disappointed in the fact that they are not getting access to the information. We deal with this by offering teachers assigned to the control group the same tools and a similar seminar the following school year, meaning that they will get the treatment one year after the teachers in the treatment group.¹⁹ In addition, the time constraints most high school teachers face also provides us with an advantage in this regard.

Third is the issue of testing. Testing can be a problem because participants become familiar with the questions and measures, and answer in the way they think we want them to or give what they believe is the "correct" answer (Creswell & Creswell, 2017). We try to address this by performing the pre-test before the exposure to the growth mindset curriculum and wait until well after the intervention is finished to perform the post-test. However, we cannot be certain that the participants don't have the intervention in mind when answering the questions in the post-test.

All of the limitations mentioned above are threats to the internal validity of the experiment. There are also threats to the external validity, that is whether it is possible to generalize the findings, and the most obvious one is selection (Cartwright, 2012; Creswell & Creswell, 2017). In our RCT, teachers self-select into treatment. Given that teachers are randomized into treatment and control group after this self-selection, we do not believe that teachers in the treatment and control group are inherently different, thus we do believe that the results we find are representative for the sample we have. However, this also means that the teachers signing up for the project might be more motivated and believe more in the message of the intervention than the average teacher, making it difficult to generalize the findings to other teachers. The problem of generalizing the result also arises when dealing with other settings and times; the students and teachers might respond differently, meaning that, as with all RCTs, we would need to perform

¹⁹However, because teachers will not always follow their students through high school, the students in the control group might not benefit from this.

several replications before making the final conclusions with regards to the overall effect of the intervention.

Most of the limitations or threats we face happen because in real life situations we are not able to control every aspect of the experiment in the same way we would in a lab experiment. However, despite this, we believe that there are many advantages in doing real life experiments as they are better at explaining what actually happens in real life, where it is impossible to control every aspect of human behavior. Thus, by testing specific interventions in a real context, we understand a little better what works, and maybe even more importantly, what doesn't.

9 Conclusion

Lifelong learning as well as scaling effective interventions that promote social and emotional competencies have been highlighted as important in previous research (Heckman, 2000). In this study we test an intervention focused on increasing the prevalence of the growth mindset, a supposedly important social and emotional competence with regards to lifelong learning (Dweck, 2009), among students in Norwegian high schools, including a component that could be beneficial for scaling the interventions; teachers. We do this by giving teachers information about the growth mindset research, in addition to tools to help them pass the information on to their students. In addition, we measure the effects on students' mindset a while after the intervention is completed.

We find that the intervention has a positive effect on students' mindset in general, as well as their mindset when it comes to math in particular. We also find that students exposed to the treatment are more likely to choose challenge, with positive effects on the number of challenging assignments chosen when given the opportunity to create their own worksheet. In addition, we find positive effects on choosing the more challenging assignment when given a hypothetical choice between a challenging and an easier assignment. Among students able to choose between different math courses in high school, we also find a positive effect on choosing the more advanced math course. Unlike the other outcomes, this is a real life choice, which could possibly influence students' career possibilities later on (Sells, 1980; Freeman & Aspray, 1999; Ma & Johnson, 2008). With regards to achievement and perseverance, we do not have sufficient power to detect the estimated effects, so we cannot conclude that the treatment had any effect on these outcomes.

When looking at the effects separately in subsamples of students, we find indications of different treatment effects based on choice of track in high school. We find that students attending the academic track seem to be affected in terms of mindset, challenge-seeking, especially choice of math course, and possibly achievement. Students attending the vocational track seem to be affected in terms of mindset (however, much less so than students attending the academic track) and challenge-seeking behavior. For students attending the vocational track, we also find indications of negative effects on achievement and perseverance (however, these are not significant). Possible concerns here are that the growth

mindset curriculum is better suited to match the interests of students attending the academic track than students attending the vocational track, or that students attending the vocational track do not really believe in the message. Our findings suggest that we might have to proceed in different ways in order to develop more of a growth mindset among students attending the vocational track compared to students attending the academic track. These are important new insights into getting more students to believe in their abilities to learn in Norwegian high schools and maybe even help reduce drop-out rates.

References

- Alan, S., Boneva, T., & Ertac, S. (2019). Ever Failed, Try Again, Succeed Better: Results from a Randomized Educational Intervention on Grit. *The Quarterly Journal of Economics*, *134*(3), 1121–1162. doi: <https://doi.org/10.1093/qje/qjz006>
- Andersen, S. C., & Nielsen, H. S. (2016). Reading intervention with a growth mindset approach improves children's skills. *Proceedings of the National Academy of Sciences*, *113*(43), 12111–12113. doi: <https://doi.org/10.1073/pnas.1607946113>
- Aronson, J., Fried, C. B., & Good, C. (2002). Reducing the Effects of Stereotype Threat on African American College Students by Shaping Theories of Intelligence. *Journal of Experimental Social Psychology*, *38*(2), 113–125. doi: <https://doi.org/10.1006/jesp.2001.1491>
- Auestad, M. L., Rege, M., & Yeager, D. (2019). Study registration - Involving the teachers: Effects of a growth mindset intervention with teacher involvement. *AEA RCT Registry*. doi: <https://doi.org/10.1257/rct.3844-1.0>
- Bettinger, E. P., Ludvigsen, S., Rege, M., Solli, I. F., & Yeager, D. (2018). Increasing perseverance in math: Evidence from a field experiment in Norway. *Journal of Economic Behavior & Organization*, *146*, 1–15. doi: <https://doi.org/10.1016/j.jebo.2017.11.032>
- Björklund, A., & Salvanes, K. G. (2011). Education and Family Background: Mechanisms and Policies. In *Handbook of the Economics of Education* (Vol. 3, pp. 201–247). The Netherlands: North-Holland, an imprint of Elsevier. doi: <https://doi.org/10.1016/B978-0-444-53429-3.00003-X>
- Blackwell, L. S., Trzesniewski, K. H., & Dweck, C. S. (2007). Implicit Theories of Intelligence Predict Achievement Across an Adolescent Transition: A Longitudinal Study and an Intervention. *Child Development*, *78*(1), 246–263. doi: <https://doi.org/10.1111/j.1467-8624.2007.00995.x>
- Burnette, J. L., O'boyle, E. H., VanEpps, E. M., Pollack, J. M., & Finkel, E. J. (2013). Mind-sets matter: A meta-analytic review of implicit theories and self-regulation. *Psychological Bulletin*, *139*(3), 655–701. doi: <https://doi.org/10.1037/a0029531>
- Cartwright, N. (2012). *RCTs, evidence, and predicting policy effectiveness*. Oxford: Oxford University Press.
- Chetty, R., Friedman, J. N., & Rockoff, J. E. (2014a). Measuring the Impacts of Teachers I: Evaluating Bias in Teacher Value-Added Estimates. *American Economic Review*, *104*(9), 2593–2632. doi: <https://doi.org/10.1257/aer.104.9.2593>
- Chetty, R., Friedman, J. N., & Rockoff, J. E. (2014b). Measuring the Impacts of Teachers II: Teacher Value-Added and Student Outcomes in Adulthood. *American Economic Review*, *104*(9), 2633–2679. doi: <https://doi.org/10.1257/aer.104.9.2633>
- Creswell, J. W., & Creswell, J. D. (2017). *Research design: Qualitative, quantitative, and mixed methods approaches*. US: SAGE publications.
- Dee, T. S. (2007). Teachers and the Gender Gaps in Student Achievement. *Journal of Human Resources*, *42*(3), 528–554. doi: <https://doi.org/10.3368/jhr.XLII.3.528>
- Deming, D. J. (2017). The Growing Importance of Social Skills in the Labor Market. *The Quarterly Journal of Economics*, *132*(4), 1593–1640. doi: <https://doi.org/10.1093/qje/qjx022>

- Dweck, C. S. (2006). *Mindset: The New Psychology of Success*. New York: The Random House Publishing Group.
- Dweck, C. S. (2009). Who will the 21st-century learners be? *Knowledge Quest*, *38*(2), 8–10.
- Dweck, C. S., & Leggett, E. L. (1988). A social-cognitive approach to motivation and personality. *Psychological Review*, *95*(2), 256–273. doi: <https://doi.org/10.1037/0033-295X.95.2.256>
- Edin, P.-A., Fredriksson, P., Nybom, M., & Ockert, B. (2017). The Rising Return to Non-cognitive Skill. *IZA Discussion Paper No. 10914*. Retrieved from <https://ssrn.com/abstract=3029784>
- Falch, T., Johannesen, A. B., & Strøm, B. (2009). *Kostnader av frafall i videregående opplæring* [Government Document]. Online. Retrieved from <https://www.regjeringen.no/globalassets/upload/kd/vedlegg/grunnskole/frafall/kostnader-av-frafall.pdf> (Accessed March 2020)
- Freeman, P., & Aspray, W. (1999). *The Supply of Information Technology Workers in the United States* (Tech. Rep.). ERIC. Retrieved from <https://files.eric.ed.gov/fulltext/ED459346.pdf>
- Good, C., Aronson, J., & Inzlicht, M. (2003). Improving adolescents' standardized test performance: An intervention to reduce the effects of stereotype threat. *Journal of Applied Developmental Psychology*, *24*(6), 645–662. doi: <https://doi.org/10.1016/j.appdev.2003.09.002>
- Grossman, P., Loeb, S., Cohen, J., & Wyckoff, J. (2013). Measure for Measure: The Relationship between Measures of Instructional Practice in Middle School English Language Arts and Teachers' Value-Added Scores. *American Journal of Education*, *119*(3), 445–470. doi: <https://doi.org/10.1086/669901>
- Gunderson, E. A., Gripshover, S. J., Romero, C., Dweck, C. S., Goldin-Meadow, S., & Levine, S. C. (2013). Parent Praise to 1- to 3-Year-Olds Predicts Children's Motivational Frameworks 5 Years Later. *Child Development*, *84*(5), 1526–1541. doi: <https://doi.org/10.1111/cdev.12064>
- Haimovitz, K., & Dweck, C. S. (2017). The Origins of Children's Growth and Fixed Mindsets: New Research and a New Proposal. *Child Development*, *88*(6), 1849–1859. doi: <https://doi.org/10.1111/cdev.12955>
- Haimovitz, K., Wormington, S. V., & Corpus, J. H. (2011). Dangerous mindsets: How beliefs about intelligence predict motivational change. *Learning and Individual Differences*, *21*(6), 747–752. doi: <https://doi.org/10.1016/j.lindif.2011.09.002>
- Hanushek, E. A. (1971). Teacher Characteristics and Gains in Student Achievement: Estimation Using Micro Data. *American Economic Review*, *61*(2), 280–288. Retrieved from <https://www.jstor.org/stable/1817003>
- Heckman, J. J. (2000). Policies to foster human capital. *Research in Economics*, *54*(1), 3–56. doi: <https://doi.org/10.1006/reec.1999.0225>
- Heckman, J. J., & Kautz, T. (2012). Hard evidence on soft skills. *Labour Economics*, *19*(4), 451–464. doi: <https://doi.org/10.1016/j.labeco.2012.05.014>

- Heckman, J. J., Stixrud, J., & Urzua, S. (2006). The Effects of Cognitive and Noncognitive Abilities on Labor Market Outcomes and Social Behavior. *Journal of Labor Economics*, *24*(3), 411–482. doi: <https://doi.org/10.1086/504455>
- Kamins, M. L., & Dweck, C. S. (1999). Person versus process praise and criticism: Implications for contingent self-worth and coping. *Developmental Psychology*, *35*(3), 835–847. doi: <https://doi.org/10.1037/0012-1649.35.3.835>
- Kane, T. J., Taylor, E. S., Tyler, J. H., & Wooten, A. L. (2011). Identifying Effective Classroom Practices Using Student Achievement Data. *Journal of Human Resources*, *46*(3), 587–613. doi: <https://doi.org/10.3368/jhr.46.3.587>
- Kautz, T., Heckman, J. J., Diris, R., Ter Weel, B., & Borghans, L. (2014). Fostering and Measuring Skills: Improving Cognitive and Non-Cognitive Skills to Promote Lifetime Success. *NBER Working Paper No. 20749*. doi: <https://doi.org/10.3386/w20749>
- Kraft, M. A. (2019). Teacher Effects on Complex Cognitive Skills and Social-Emotional Competencies. *Journal of Human Resources*, *54*(1), 1–36. doi: <https://doi.org/10.3368/jhr.54.1.0916.8265R3>
- Lavecchia, A. M., Liu, H., & Oreopoulos, P. (2016). Behavioral Economics of Education: Progress and Possibilities. In *Handbook of the Economics of Education* (Vol. 5, pp. 1–74). The Netherlands: North-Holland, an imprint of Elsevier. doi: <https://doi.org/10.1016/B978-0-444-63459-7.00001-4>
- Ma, X., & Johnson, W. (2008). Mathematics as the critical filter: Curricular effects on gendered career choices. In *In H. M. G. Watt & J. S. Eccles (Eds.), Gender and occupational outcomes: Longitudinal assessments of individual, social, and cultural influences* (pp. 55–83). Washington, DC: American Psychological Association. doi: <https://doi.org/10.1037/11706-002>
- Markussen, E., & Gloppen, S. K. (2012). *Påbygg – et gode eller en nødløsning? En studie av påbygging til generell studiekompetanse i Østfold, Akershus, Buskerud, Rogaland og Nord-Trøndelag skoleåret 2010-2011* (Tech. Rep.). Retrieved from <https://nifu.brage.unit.no/nifu-xmlui/bitstream/handle/11250/280848/NIFUrapport2012-2.pdf?sequence=1&isAllowed=y>
- Mueller, C. M., & Dweck, C. S. (1998). Praise for Intelligence Can Undermine Children's Motivation and Performance. *Journal of Personality and Social Psychology*, *75*(1), 33–52. doi: <https://doi.org/10.1037/0022-3514.75.1.33>
- Murnane, R. J. (1975). *The Impact of School Resources on the Learning of Inner City Children*. Cambridge: Ballinger Pub. Co.
- OECD Publishing. (2015). *Skills for Social Progress: The Power of Social and Emotional Skills*. Paris: OECD Publishing. (OECD Skills Studies) doi: <https://doi.org/10.1787/9789264226159-en>
- Paunesku, D., Walton, G. M., Romero, C., Smith, E. N., Yeager, D. S., & Dweck, C. S. (2015). Mind-Set Interventions Are a Scalable Treatment for Academic Underachievement. *Psychological Science*, *26*(6), 784–793. doi: <https://doi.org/10.1177/0956797615571017>
- Petersen, J. L., & Hyde, J. S. (2014). Gender-Related Academic and Occupational Interests and Goals. In *Advances in Child Development and Behavior* (Vol. 47, pp.

- 43–76). US/UK: Academic Press, an imprint of Elsevier. doi: <https://doi.org/10.1016/bs.acdb.2014.04.004>
- Pomerantz, E. M., & Kempner, S. G. (2013). Mothers' daily person and process praise: Implications for children's theory of intelligence and motivation. *Developmental Psychology*, *49*(11), 2040–2046. doi: <https://doi.org/10.1037/a0031840>
- Rattan, A., Savani, K., Chugh, D., & Dweck, C. S. (2015). Leveraging Mindsets to Promote Academic Achievement: Policy Recommendations. *Perspectives on Psychological Science*, *10*(6), 721–726. doi: <https://doi.org/10.1177/1745691615599383>
- Rege, M., Størksen, I., Solli, I. F., Kalil, A., McClelland, M., Braak, D. t., ... others (2019). Promoting Child Development in a Universal Preschool System: A Field Experiment. Retrieved from <https://ssrn.com/abstract=3434830>
- Rockoff, J. E. (2004). The Impact of Individual Teachers on Student Achievement: Evidence from Panel Data. *American Economic Review*, *94*(2), 247–252. doi: <https://doi.org/10.1257/0002828041302244>
- Sells, L. W. (1980). Mathematics: The Invisible Filter. *Engineering Education*, *70*(4), 340–341. Retrieved from <https://eric.ed.gov/?id=EJ216630>
- Statistics Norway. (2019). *Gjennomføring i videregående opplæring*. Retrieved from <https://www.ssb.no/utdanning/statistikker/vgogjen> (Accessed June 2019)
- Walton, G. M., & Yeager, D. S. (in press). Seed and Soil: Psychological Affordances in Contexts Help to Explain Where Wise Interventions Succeed or Fail. *Current Directions in Psychological Science*.
- Yeager, D. S., & Dweck, C. S. (2012). Mindsets That Promote Resilience: When Students Believe That Personal Characteristics Can Be Developed. *Educational Psychologist*, *47*(4), 302–314. doi: <https://doi.org/10.1080/00461520.2012.722805>
- Yeager, D. S., Hanselman, P., Walton, G. M., Murray, J. S., Crosnoe, R., Muller, C., ... others (2019). A national experiment reveals where a growth mindset improves achievement. *Nature*, *573*(7774), 364–369.
- Yeager, D. S., Johnson, R., Spitzer, B. J., Trzesniewski, K. H., Powers, J., & Dweck, C. S. (2014). The far-reaching effects of believing people can change: Implicit theories of personality shape stress, health, and achievement during adolescence. *Journal of personality and social psychology*, *106*(6), 867–884. doi: <https://doi.org/10.1037/a0036335>
- Yeager, D. S., Romero, C., Paunesku, D., Hulleman, C. S., Schneider, B., Hinojosa, C., ... others (2016). Using design thinking to improve psychological interventions: The case of the growth mindset during the transition to high school. *Journal of Educational Psychology*, *108*(3), 374–391. doi: <https://doi.org/10.1037/edu0000098>
- Yeager, D. S., & Walton, G. M. (2011). Social-psychological interventions in education: They're not magic. *Review of Educational Research*, *81*(2), 267–301. doi: <https://doi.org/10.3102/0034654311405999>
- Yeager, D. S., Walton, G. M., Brady, S. T., Akcinar, E. N., Paunesku, D., Keane, L., ... others (2016). Teaching a lay theory before college narrows achievement gaps at scale. *Proceedings of the National Academy of Sciences*, *113*(24), 3341–3348. doi: <https://doi.org/10.1073/pnas.1524360113>

Table 1: Descriptive statistics and balance test

	Consent			Survey 1			Survey 2		
	Control	Treat	Diff	Control	Treat	Diff	Control	Treat	Diff
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Vocational track	0.49 (0.50)	0.39 (0.49)	0.000 (.)	0.49 (0.50)	0.39 (0.49)	0.000 (.)	0.49 (0.50)	0.39 (0.49)	0.000 (.)
Baseline mindset	0.01 (1.00)	-0.01 (1.00)	-0.069 (0.057)	0.01 (1.00)	-0.01 (1.00)	-0.069 (0.057)	0.02 (1.00)	-0.02 (1.00)	-0.108 (0.068)
Baseline math mindset	0.03 (1.00)	-0.03 (1.00)	-0.067 (0.051)	0.03 (1.00)	-0.03 (1.00)	-0.067 (0.051)	0.02 (1.02)	-0.02 (0.99)	-0.041 (0.068)
10th grade GPA	-0.04 (1.02)	0.04 (0.98)	0.000 (0.054)	-0.04 (1.01)	0.05 (0.96)	0.019 (0.055)	-0.05 (1.02)	0.05 (0.98)	0.022 (0.063)
10th grade math	-0.02 (1.01)	0.02 (0.99)	-0.012 (0.055)	-0.01 (1.02)	0.02 (0.99)	-0.016 (0.056)	-0.03 (1.01)	0.03 (0.99)	0.045 (0.074)
Female	0.46 (0.50)	0.51 (0.50)	0.006 (0.026)	0.46 (0.50)	0.51 (0.50)	0.001 (0.027)	0.48 (0.50)	0.54 (0.50)	0.018 (0.031)
Immigrant	0.28 (0.45)	0.28 (0.45)	-0.026 (0.031)	0.28 (0.45)	0.28 (0.45)	-0.026 (0.031)	0.28 (0.45)	0.26 (0.44)	-0.056 (0.039)
Older than 16	0.10 (0.30)	0.08 (0.27)	-0.020+ (0.011)	0.09 (0.29)	0.08 (0.26)	-0.028* (0.012)	0.10 (0.30)	0.08 (0.26)	-0.030* (0.013)
<i>p</i> -value on joint significance			0.585			0.599			0.575
Observations	1 194	1 355	2 549	1 090	1 274	2 364	850	918	1 768

Note: For each sample - "Consent", "Survey 1" and "Survey 2" - the columns provide the mean and standard deviations for students in the control and treatment groups. Estimated coefficients with robust standard errors clustered at class level, as well as educational program and school fixed effects, from regressing against treatment status are reported in the columns labeled "Diff" for the different samples.

+ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$.

Table 2: Treatment effects on students' mindset

	Mindset		Math mindset	
	(1)	(2)	(3)	(4)
Panel A: full sample				
Treatment	0.167** (0.040)	0.202** (0.039)	0.198** (0.050)	0.205** (0.051)
Observations	1 757	1 757	1 751	1 751
Adjusted R^2	0.036	0.176	0.030	0.168
Panel B: vocational track				
Treatment	0.079 (0.048)	0.106* (0.048)	0.128 (0.077)	0.142* (0.070)
Observations	765	765	766	766
Adjusted R^2	0.027	0.135	0.023	0.165
Panel C: academic track				
Treatment	0.262** (0.061)	0.307** (0.060)	0.273** (0.062)	0.264** (0.074)
Observations	992	992	985	985
Adjusted R^2	0.041	0.205	0.031	0.167
Panel D: vocational track by GPA				
Treatment	0.083 (0.088)	0.112 (0.087)	0.085 (0.127)	0.081 (0.121)
High \times treatment	-0.026 (0.139)	0.001 (0.137)	0.069 (0.172)	0.094 (0.156)
Observations	765	765	766	766
Adjusted R^2	0.049	0.137	0.036	0.178
Panel E: academic track by GPA				
Treatment	0.340** (0.107)	0.425** (0.097)	0.317** (0.104)	0.367** (0.112)
High \times treatment	-0.147 (0.130)	-0.201+ (0.104)	-0.080 (0.118)	-0.168 (0.116)
Observations	992	992	985	985
Adjusted R^2	0.055	0.206	0.033	0.164
Control variables		X		X

Note: The first row lists the dependent variable. Panels represent different samples, with the full sample in Panel A, vocational track students in Panels B and D, and academic track students in Panels C and E. Each column and row presents a separate regression and reports the estimated treatment coefficients with robust standard errors clustered at class level, number of observations and adjusted R^2 . Fixed effects for school and educational program are included in all specifications. The dependent variables are standardized mindset in columns 1 and 2 and standardized math mindset in columns 3 and 4. The independent variables of interest are the treatment in Panels A, B and C as well as the interaction between treatment and high GPA in Panels D and E. Survey 2 sample (n=1 768) (11 students have not answered any of the two general mindset questions and 17 students have not answered the math mindset question in survey 2).

+ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$.

Table 3: Treatment effects on challenge-seeking

	Challenge		Difficult task		Advanced math ¹	
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: full sample						
Treatment	0.075 (0.046)	0.082+ (0.044)	0.046+ (0.024)	0.052* (0.023)	0.067 (0.050)	0.057* (0.027)
Observations	1 669	1 669	1 729	1 729	994	994
Adjusted R^2	0.034	0.056	0.052	0.153	0.169	0.395
Panel B: vocational track						
Treatment	0.102 (0.068)	0.116+ (0.068)	0.054+ (0.032)	0.055+ (0.032)	N/A	N/A
Observations	744	744	757	757		
Adjusted R^2	0.073	0.075	0.046	0.126		
Panel C: academic track						
Treatment	0.045 (0.062)	0.055 (0.052)	0.038 (0.037)	0.042 (0.033)	0.067 (0.050)	0.057* (0.027)
Observations	925	925	972	972	994	994
Adjusted R^2	0.004	0.047	0.038	0.156	0.169	0.395
Panel D: vocational track by GPA						
Treatment	0.135 (0.110)	0.141 (0.106)	0.033 (0.054)	0.035 (0.053)		
High \times treatment	-0.063 (0.176)	-0.056 (0.174)	0.033 (0.078)	0.046 (0.075)	N/A	N/A
Observations	744	744	757	757		
Adjusted R^2	0.072	0.064	0.059	0.123		
Panel E: academic track by GPA						
Treatment	0.044 (0.099)	0.028 (0.104)	-0.004 (0.051)	-0.017 (0.050)	0.051 (0.059)	0.048 (0.045)
High \times treatment	-0.002 (0.144)	0.026 (0.156)	0.069 (0.066)	0.079 (0.063)	0.014 (0.063)	0.010 (0.056)
Observations	925	925	972	972	994	994
Adjusted R^2	0.003	0.041	0.063	0.165	0.247	0.404
Control variables		X		X		X

Note: The first row lists the dependent variable. Panels represent different samples, with the full sample in Panel A, vocational track students in Panels B and D, and academic track students in Panels C and E. Each column and row presents a separate regression and reports the estimated treatment coefficients with robust standard errors clustered at class level, number of observations and adjusted R^2 . Fixed effects for school and educational program are included in all specifications. The dependent variables are standardized number of challenging assignments chosen in columns 1 and 2, an indicator for choosing the more difficult task if given the choice in columns 3 and 4 and an indicator for choosing the advanced math course in columns 5 and 6. The independent variables of interest are the treatment in Panels A, B and C as well as the interaction between treatment and high GPA in Panels D and E. Survey 2 sample (n=1 768) (99 students have not chosen any assignments in the create your own worksheet task and 39 did not choose between the challenging and the easy assignment in survey 2). ¹In columns 5 and 6 the sample is restricted to only those attending the academic track, as these are the students that have the option of choosing the advanced math course in high school, hence there are no observations in the vocational track subsample (5 students were missing the math code for the first year in high school).

+ p < 0.10, * p < 0.05, ** p < 0.01.

Table 4: Treatment effects on achievement

	Math quiz		GPA		Completed	
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: full sample						
Treatment	0.025 (0.061)	0.006 (0.052)	0.067 (0.061)	0.030 (0.040)	-0.003 (0.012)	-0.010 (0.011)
Observations	1 710	1 710	1 768	1 768	1 768	1 768
Adjusted R^2	0.196	0.346	0.135	0.491	0.045	0.132
Panel B: vocational track						
Treatment	-0.025 (0.064)	-0.037 (0.057)	0.073 (0.088)	-0.004 (0.065)	-0.003 (0.017)	-0.011 (0.016)
Observations	753	753	769	769	769	769
Adjusted R^2	0.085	0.218	0.119	0.456	0.053	0.145
Panel C: academic track						
Treatment	0.080 (0.108)	0.073 (0.087)	0.061 (0.086)	0.051 (0.048)	-0.002 (0.018)	-0.008 (0.017)
Observations	957	957	999	999	999	999
Adjusted R^2	0.156	0.350	0.132	0.534	0.033	0.130
Panel D: vocational track by GPA						
Treatment	-0.082 (0.084)	-0.072 (0.080)	-0.012 (0.125)	-0.097 (0.122)	0.003 (0.031)	-0.007 (0.030)
High \times treatment	0.103 (0.149)	0.097 (0.132)	0.103 (0.157)	0.167 (0.139)	-0.019 (0.040)	-0.008 (0.039)
Observations	753	753	769	769	769	769
Adjusted R^2	0.104	0.227	0.376	0.520	0.080	0.145
Panel E: academic track by GPA						
Treatment	0.155 (0.130)	0.157 (0.115)	-0.073 (0.092)	-0.028 (0.066)	-0.050 (0.032)	-0.051+ (0.028)
High \times treatment	-0.148 (0.155)	-0.194 (0.141)	0.201+ (0.113)	0.118 (0.088)	0.081* (0.032)	0.077** (0.028)
Observations	957	957	999	999	999	999
Adjusted R^2	0.218	0.372	0.404	0.567	0.066	0.156
Control variables		X		X		X

Note: The first row lists the dependent variable. Panels represent different samples, with the full sample in Panel A, vocational track students in Panels B and D, and academic track students in Panels C and E. Each column and row presents a separate regression and reports the estimated treatment coefficients with robust standard errors clustered at class level, number of observations and adjusted R^2 . Fixed effects for school and educational program are included in all specifications. The dependent variables are standardized number of correct answers on the math quiz at the end of survey 2 in columns 1 and 2, standardized GPA in columns 3 and 4, and an indicator for completed first year in high school in columns 5 and 6. The independent variables of interest are the treatment in Panels A, B and C as well as the interaction between treatment and high GPA in Panels D and E. Survey 2 sample (n=1 768) (58 students were not able to get to the math quiz in survey 2, hence no answers are registered for these students).

+ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$.

Appendix A Correlations

In [Table A1](#), we present a correlation matrix of our pre-treatment mindset measures, grades and track in high school. For this correlation matrix we use the sample from the pre-intervention survey, "Survey 1", with 2 364 students.²⁰ Our measure of students' baseline mindset is based on students' answers to the survey questions "Mindset 1" and "Mindset 2", thus these are highly correlated with the baseline mindset measure. "Mindset 1" and "Mindset 2" are also highly correlated with each other, which is no surprise as they represent two different ways of asking the same question. With regards to the math mindset measure, there is still somewhat of a correlation between this and the baseline measure and "Mindset 1" and "Mindset 2", albeit not as strong. This also makes sense as the math mindset measure is posed in a different way and relates only to one subject - math. In addition, similar to previous studies (Good et al., 2003; Paunesku et al., 2015; Bettinger et al., 2018), we also see that there is a positive correlation between grades (both GPA and math) and the mindset measures (of between 0.15 and 0.20). With regards to track in high school, there is a negative correlation between the mindset measures and attending the vocational track of between 0.10 and 0.15, as well as a negative correlation between the vocational track and grades of around 0.40.

Table A1: Correlation between pre-treatment mindset measures, grades and track in high school

	Baseline mindset	Math mindset	Mindset		10th grade	
			1	2	GPA	math
Math mindset	0.421**					
Mindset 1	0.917**	0.382**				
Mindset 2	0.917**	0.386**	0.682**			
GPA 10th grade	0.194**	0.156**	0.169**	0.186**		
Math grade 10th grade	0.166**	0.159**	0.147**	0.156**	0.764**	
Vocational track	-0.145**	-0.117**	-0.130**	-0.134**	-0.410**	-0.370**

Note: Survey 1 sample (n=2 364)
+ p < 0.10, * p < 0.05, ** p < 0.01.

²⁰In [Online Appendix B](#) we also include cross tabulations of pre-intervention mindset, as well as track in high school and GPA.

Appendix B Questions about challenge-seeking behavior

"Make your own worksheet" task:

"We are interested in which type of math assignment first year high school students prefer to work on. You are now given the opportunity to make your own worksheet. At the end of the session you will have the opportunity to answer some of these assignments. On the next few pages there are problems from three different areas in mathematics. Choose between two and six assignments from each area. You can choose from assignments that are:

Very challenging - but you might learn a lot

Somewhat challenging - and you might learn a medium amount

Not very challenging - and you probably won't learn very much

Do not try to answer the math problems. Just click the problems you'd like to try later if there's time."

The areas from which the students are given the opportunity to choose assignments are numbers and algebra, measurement and geometry, and statistics and probability.

Hypothetical scenario:

"Imagine that, tomorrow, your math teacher hands out two hand-in assignments. You get to choose which one to do. You're grade will not be affected by the assignment. One choice is an easy review: it has math problems you already know how to solve, and you will probably get most of the answers right without having to think very much. It takes 30 minutes to solve. The other choice is a hard challenge: it has math problems you don't know how to solve, and you will probably get most of the problems wrong, but you might learn something new. It also takes 30 minutes. If you had to choose right now, which one would you choose?"

Easy assignment: "The easy math assignment where I would get most problems right."

Challenging assignment: "The hard math assignment where I would possibly learn something new."

Appendix C Math quiz

In order to have comparable short-term achievement measures for the students in the treatment and control groups, we included a math quiz towards the end of the post-intervention survey. Given the amount of questions the students had to answer in this round, and the time constraints teachers in high schools experience,²¹ we chose 10 of the multiple choice algebra questions posed to the students in Bettinger et al. (2018).²² Prior to the math quiz, the students were given the following information:

"On the next few pages we have chosen ten math assignments for you to work on. Some of the assignments will be fairly easy, while others could be pretty difficult. This is not a test, and how you answer will not affect your grades. Your answers will remain confidential, and we are not going to share them with anyone at your school. This means that neither your peers, your teacher or your parents will know how you perform. In order to solve the assignments, you should have a pencil and something to write on to hand. The assignments should be solved individually."

The algebra questions were somewhat challenging for most students, and only 6 students managed to get all of them correct. On average, students got 3 questions right (see [Figure OC.7](#) in Online Appendix C).

One important feature of the test situation was that neither the teachers nor the students knew that they would be answering algebra questions towards the end of the post-intervention survey, so there was no way for them to prepare. In fact, many teachers complained about this afterwards, since they did not get a chance to prepare their students. But there were also teachers who saw it as an opportunity to get the students to use what they had learned about the growth mindset. One of the teachers in the treatment group actually reported back that this was a nice opportunity to really get to use what they had been practicing and motivate their students with regards to the growth mindset.

²¹Making it very difficult to conduct a third set of questions for all the students in our sample.

²²The reasons for selecting 10 questions was that this is the amount of questions most students managed to get through in their study (Bettinger et al., 2018).

Appendix D Questions about educational attitudes and expectations

Students' attitudes to getting a bad score on a math test are measured by asking them to:

"Imagine that you get a bad grade on a test tomorrow, how likely are you to think that:"

Attribution: "This shows that I am not very good in math."

Strategy: "I can do better next time if I figure out a better way to work."

which they are to rate on a five-point scale ranging from "Unlikely" to "Very likely". All possible answers being: "Unlikely", "Not likely", "Quite likely", "Likely" and "Very likely".

In addition, the students are asked to which degree, on the same six-point scale as used to map students' mindset, they agree with the following statements:

Anxious: "I become insecure and anxious when thinking about high school math."

Interest: "High school math is going to be interesting."

and to rate:

Expectation: "How do you think you will perform in high school math?"

on a six-point scale ranging from "Very bad" to "Very good". With all possible answers being: "Very bad", "Bad", "Somewhat bad", "Somewhat good", "Good" and "Very good".

Appendix E Missing observations for individual characteristics

We do experience some missing observations for individual characteristics, both in the survey data and in the data from the government administrative records. However, we want to make use of the full sample for the students in which we have outcome measures, and therefore impute the individual characteristics for those that were missing. We do this by re-coding the missing data points for the individual characteristics and generating a dummy variable equal to one if the observation is initially missing.

The individual characteristics where we have missing observations are presented in [Table D1](#) below, and include baseline mindset, baseline math mindset, math grade and GPA from middle school, gender (indicator for female) and immigrant. For all missing data points, we re-code the missing observation as the mean of that variable in the student's class. In the rare event that the variable is missing for everyone in the class, we use the mean of that variable in the student's school or the entire sample.²³ In the regressions, we use the variable without any missing observations and also include the indicator for missing on the different characteristics.

Table D1: Missing observations for baseline individual characteristics

	Observed characteristics	Missing	Percent missing
Baseline mindset	1 579	189	10.69
Baseline math mindset	1 576	192	10.86
10th grade GPA	1 718	50	2.83
10th grade math grade	1 669	99	5.60
Female	1 753	15	0.85
Immigrant	1 555	213	12.05
Full sample	1 768		

²³Applies to 25 students in total on the characteristics for baseline mindset, baseline math mindset and indicator for immigrant.

Appendix F Attrition

We do experience a high attrition rate in our study, with 781 students, or about 30 percent of our "Consent" sample not being present in our "Survey 2" sample (ref [Section 5](#)). In [Section 6](#) we mention three possible sources of this attrition; (1) teachers who withdrew from the project, taking with them entire classes of students; (2) absence and drop-out, preventing students from being present at school when the surveys were conducted; and (3) that students performing the surveys skip some of the questions in the survey.

Most of the attrition can be attributed to the first of these; teachers withdrawing from the project. Of the 205 teachers who signed up for the research project, about 25 percent withdrew because of time constraints, sick leave, quitting or other reasons. In the post-intervention survey therefore, 127 classes in 29 different schools and 72 blocks participated. Our biggest concern with teachers dropping out is that we are left with an unbalanced sample of classes within schools. This can happen if more teachers assigned to the treatment group withdraw their classes compared to teachers in the control group (or vice versa) or if different kinds of teachers assigned to the treatment group drop out compared to teachers assigned to the control group (that is, if only one class in each block drops out). Of the classes that dropped out, 27 were in the treatment group and 26 were in the control group during the course of the project. 17 entire blocks and all participating classes at 3 schools dropped out. Based on these numbers, we are not too concerned with unbalance in who drops out, since we are losing the same number of classes in both the treatment and control group. In addition, since we lost 17 entire blocks, teachers seem to also withdraw in pairs (as they did when signing up), which is beneficial for our design. However, below we include several attrition analyses in order to make sure that we are losing the same students in the control and treatment group.

We do also experience some attrition due to the second source; absence and drop-out. In the "Consent" sample, consisting of all students who entered either the pre- or the post-intervention survey, we have on average 16 students per class, whereas in "Survey 2", the average number of students per class is 14, indicating that about 2 students in each class were absent when performing the post-intervention survey. As a consequence, we could end up with an unbalanced sample if students dropping out from the treatment group are different kinds of students than those leaving the control group (as above). With regard to the last source, that students skip some questions in the survey, this is only a minor problem, with at most 99 students not answering (the "make your own worksheet" task).

In order to figure out if we are losing the same kind of students in the treatment and the control group, we perform several attrition analyses to investigate if attrition might depend on treatment status or individual characteristics. First, in [Table E1](#) we investigate attrition in relation to the outcome variables. We do this by first regressing the treatment on not having answered any of the survey questions in the post-intervention survey or "Survey 2", before we look at attrition on each specific outcome in particular. We find

a positive estimate for attrition in "Survey 2" of 5.1 percent, suggesting that treated students are more likely not to have answered the post-intervention survey. The estimate, however, is not significant. The only significant estimate with regards to attrition is a positive effect of 1.6 percent on the advanced math outcome, suggesting different attrition based on treatment status. However, when looking at this many different outcomes, we might expect to find one significant coefficient.

Next, in [Table E2](#), we present descriptive statistics for stayers and leavers, and investigate attrition in "Survey 2", that is not having answered the post-intervention survey, against the individual characteristics. Doing this, we find that the lower the baseline math and GPA grade, the more likely it is that students have not answered the post-intervention survey. We also find that students lagging behind (i.e. are older than 16) are more likely not to have answered this second round of questions (however, this is only significant at the 10 percent level). This is in line with what we expect, since higher achieving students and students who are not lagging behind might be more likely to show up at school. In [Table E3](#), we investigate this even further and look at whether the treatment status combined with individual characteristics might affect attrition as well. Here we find indications that treated males are more likely to drop out with a significant estimate (at the 10 percent level) of 7.5 percent.

Table E1: Attrition analyses: consent sample (n=2 549)

	Observations	Missing observations	Test of attrition
	(1)	(2)	(3)
Survey 2	1 768	781	0.051 (0.037)
Mindset	1 757	792	0.047 (0.037)
Math mindset	1 751	798	0.045 (0.037)
Challenging	1 669	880	0.041 (0.036)
Difficult task	1 729	820	0.049 (0.036)
Advanced math	2 495	54	0.016** (0.005)
Number correct	1 710	839	0.049 (0.037)
GPA	2 549	0	0.000 (.)
Attribution	1 730	819	0.041 (0.037)
Strategy	1 728	821	0.041 (0.037)
Expectation	1 704	845	0.051 (0.037)
Interest	1 704	845	0.050 (0.037)
Anxious	1 702	847	0.050 (0.037)

Note: Analyses of attrition with number of observations for the different outcomes in column 1 and number of missing observations for the different outcomes in column 2. Test of attrition is presented in column 3. Each row represents a separate regression, with robust standard errors clustered at class level, as well as educational program and school fixed effects. The dependent variable is attrition on the different outcomes and the independent variable of interest is the treatment.

+ p < 0.10, * p < 0.05, ** p < 0.01.

Table E2: Attrition survey 2: consent sample (n=2 549)

	Stayers (1)	Leavers (2)	Difference (3)
Vocational track	0.43 (0.50)	0.44 (0.50)	0.000 (.)
Baseline mindset	0.01 (0.99)	-0.03 (1.02)	-0.007 (0.009)
Baseline math mindset	0.02 (0.99)	-0.04 (1.03)	-0.006 (0.008)
10th grade GPA	0.03 (0.97)	-0.07 (1.07)	-0.036** (0.011)
10th grade math grade	0.02 (0.99)	-0.06 (1.03)	-0.025** (0.009)
Female	0.51 (0.50)	0.44 (0.50)	-0.016 (0.017)
Immigrant	0.27 (0.44)	0.31 (0.46)	0.016 (0.019)
Older than 16	0.09 (0.28)	0.09 (0.29)	0.055+ (0.031)
Mindset 1	0.00 (1.00)	-0.00 (1.00)	-0.001 (0.009)
Mindset 2	0.02 (0.99)	-0.04 (1.02)	-0.010 (0.008)
Observations	1 768	781	2 549

Note: Descriptive statistics for stayers and leavers are reported in columns 1 and 2, respectively. Column 3 presents the estimated coefficients with robust standard errors clustered at class level, as well as educational program and school fixed effects, from regressing each individual characteristic against attrition in Survey 2.

+ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$.

Table E3: Attrition survey 2: consent sample (n=2 549)

	Stayers			Leavers			Differences		
	Control	Treat		Control	Treat		X	Treat	Interaction
	(1)	(2)	(3)	(4)	(5)	(6)	(7)		
Vocational track	0.49 (0.50)	0.39 (0.49)	0.49 (0.50)	0.40 (0.49)	0.000 (.)	0.009 (0.061)	0.083 (0.074)		
Baseline mindset	0.03 (0.99)	-0.00 (0.99)	-0.03 (1.03)	-0.02 (1.01)	-0.010 (0.014)	0.043 (0.037)	0.006 (0.018)		
Baseline math mindset	0.04 (1.00)	-0.00 (0.97)	0.00 (1.00)	-0.08 (1.05)	0.003 (0.010)	0.042 (0.037)	-0.017 (0.016)		
10th grade GPA	-0.02 (0.99)	0.08 (0.95)	-0.10 (1.08)	-0.05 (1.05)	-0.024 (0.015)	0.051 (0.038)	-0.022 (0.019)		
10th grade math grade	-0.01 (1.00)	0.05 (0.98)	-0.05 (1.05)	-0.06 (1.01)	-0.009 (0.015)	0.058 (0.039)	-0.029 (0.018)		
Female	0.48 (0.50)	0.54 (0.50)	0.42 (0.49)	0.45 (0.50)	0.010 (0.023)	0.075+ (0.045)	-0.049 (0.034)		
Immigrant	0.28 (0.45)	0.26 (0.44)	0.29 (0.45)	0.33 (0.47)	0.024 (0.028)	0.044 (0.037)	-0.012 (0.040)		
Older than 16	0.10 (0.30)	0.08 (0.26)	0.09 (0.29)	0.10 (0.30)	0.009 (0.041)	0.044 (0.038)	0.096 (0.058)		
Mindset 1	0.01 (1.00)	-0.01 (1.00)	-0.01 (1.02)	0.00 (0.99)	-0.002 (0.013)	0.042 (0.037)	0.004 (0.016)		
Mindset 2	0.04 (1.00)	-0.00 (0.98)	-0.06 (1.04)	-0.03 (1.01)	-0.014 (0.013)	0.042 (0.037)	0.009 (0.017)		
Observations	850	918	344	437		2 549			

Note: Descriptive statistics for stayers in the control and treatment groups and leavers in the control and treatment groups are reported in columns 1 to 4, respectively. Columns 5 to 7 present the estimated coefficients with robust standard errors clustered at class level, as well as educational program and school fixed effects, from regressing each individual characteristic as well as treatment status and the interaction between the individual characteristic and treatment status against attrition in Survey 2. + p < 0.10, * p < 0.05, ** p < 0.01.

Appendix G Tables secondary outcomes

Table G1: Treatment effects on perseverance

	1st question		First 5 questions		All questions	
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: full sample						
Treatment	-0.042 (0.081)	-0.050 (0.073)	-0.055 (0.077)	-0.059 (0.073)	-0.040 (0.080)	-0.046 (0.075)
Observations	1 710	1 710	1 710	1 710	1 710	1 710
Adjusted R^2	0.170	0.256	0.168	0.245	0.183	0.272
Panel B: vocational track						
Treatment	-0.113 (0.114)	-0.101 (0.094)	-0.141 (0.106)	-0.128 (0.090)	-0.143 (0.107)	-0.134 (0.090)
Observations	753	753	753	753	753	753
Adjusted R^2	0.145	0.254	0.157	0.259	0.180	0.296
Panel C: academic track						
Treatment	0.037 (0.116)	0.041 (0.107)	0.040 (0.114)	0.045 (0.109)	0.073 (0.120)	0.074 (0.115)
Observations	957	957	957	957	957	957
Adjusted R^2	0.174	0.258	0.173	0.243	0.181	0.258
Panel D: vocational track by GPA						
Treatment	-0.121 (0.127)	-0.070 (0.110)	-0.180 (0.115)	-0.138 (0.096)	-0.196+ (0.109)	-0.159+ (0.092)
High \times treatment	-0.006 (0.174)	-0.019 (0.163)	0.061 (0.150)	0.051 (0.136)	0.086 (0.149)	0.070 (0.135)
Observations	753	753	753	753	753	753
Adjusted R^2	0.183	0.283	0.192	0.284	0.222	0.315
Panel E: academic track by GPA						
Treatment	0.092 (0.160)	0.123 (0.149)	0.103 (0.155)	0.142 (0.151)	0.160 (0.162)	0.192 (0.158)
High \times treatment	-0.106 (0.161)	-0.139 (0.147)	-0.120 (0.141)	-0.160 (0.136)	-0.161 (0.143)	-0.202 (0.137)
Observations	957	957	957	957	957	957
Adjusted R^2	0.191	0.264	0.197	0.242	0.208	0.258
Control variables		X		X		X

Note: The first row lists the dependent variable. Panels represent different samples, with the full sample in Panel A, vocational track students in Panels B and D, and academic track students in Panels C and E. Each column and row presents a separate regression and reports the estimated treatment coefficients with robust standard errors clustered at class level, number of observations and adjusted R^2 . Fixed effects for school and educational program are included in all specifications. The dependent variables are the log of time spent on the first question of the math quiz in columns 1 and 2, the first 5 questions in columns 3 and 4 and all 10 questions in columns 5 and 6. The independent variables of interest are the treatment in Panels A, B and C as well as the interaction between treatment and high GPA in Panels D and E. Survey 2 sample (n=1 768) (58 students were not able to get to the math quiz in survey 2, hence no times are recorded for these students).

+ p < 0.10, * p < 0.05, ** p < 0.01.

Table G2: Treatment effects on students' attitudes and expectations

	Attribution (1)	Strategy (2)	Expectation (3)	Interest (4)	Anxious (5)
Panel A: full sample					
Treatment	-0.027 (0.040)	0.060 (0.055)	0.052 (0.055)	0.096+ (0.051)	0.032 (0.053)
Observations	1730	1728	1704	1704	1702
Adjusted R^2	0.141	0.057	0.197	0.153	0.131
Panel B: vocational track					
Treatment	-0.014 (0.057)	0.086 (0.072)	0.036 (0.082)	0.167+ (0.084)	0.030 (0.079)
Observations	758	757	748	749	748
Adjusted R^2	0.181	0.067	0.266	0.173	0.169
Panel C: academic track					
Treatment	-0.032 (0.057)	0.038 (0.085)	0.066 (0.066)	0.029 (0.064)	0.046 (0.069)
Observations	972	971	956	955	954
Adjusted R^2	0.102	0.037	0.156	0.131	0.106
Panel D: vocational track by GPA					
Treatment	-0.084 (0.089)	0.132 (0.122)	-0.097 (0.104)	0.116 (0.113)	0.014 (0.110)
High \times treatment	0.139 (0.138)	-0.070 (0.166)	0.250 (0.163)	0.092 (0.153)	0.052 (0.159)
Observations	758	757	748	749	748
Adjusted R^2	0.174	0.059	0.265	0.168	0.176
Panel E: academic track by GPA					
Treatment	-0.115 (0.082)	0.028 (0.112)	0.165+ (0.096)	-0.056 (0.093)	-0.083 (0.109)
High \times treatment	0.163 (0.121)	-0.005 (0.122)	-0.186 (0.117)	0.121 (0.130)	0.213 (0.157)
Observations	972	971	956	955	954
Adjusted R^2	0.104	0.040	0.159	0.136	0.109
Control variables	X	X	X	X	X

Note: The first row lists the dependent variable. Panels represent different samples, with the full sample in Panel A, vocational track students in Panels B and D, and academic track students in Panels C and E. Each column and row presents a separate regression and reports the estimated treatment coefficients with robust standard errors clustered at class level, number of observations and adjusted R^2 . Fixed effects for school and educational program are included in all specifications. The dependent variables are attribution, new strategy, expectations, interest and insecure and anxious in columns 1 to 5, respectively. All dependent variables have been standardized. The independent variables of interest are the treatment in Panels A, B and C as well as the interaction between treatment and high GPA in Panels D and E. Survey 2 sample (n=1768) (38 students have not answered the attribution and 40 have not answered the strategy question, 64 have not answered the expectations and the interest question and 66 students have not answered the anxious question in survey 2).

+ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$.

Online Appendix A Excluding students without a middle school GPA from the subsample analyses

Table OA.1: Subsample analyses excluding missing: students' mindset

	Mindset		Math mindset	
	(1)	(2)	(3)	(4)
Panel A: vocational track by GPA				
Treatment	0.077 (0.086)	0.106 (0.085)	0.098 (0.136)	0.092 (0.130)
High \times treatment	-0.025 (0.135)	0.002 (0.133)	0.053 (0.182)	0.080 (0.163)
Observations	733	733	734	734
Adjusted R^2	0.049	0.142	0.038	0.187
Panel B: academic track by GPA				
Treatment	0.341** (0.111)	0.420** (0.101)	0.327** (0.108)	0.375** (0.112)
High \times treatment	-0.150 (0.133)	-0.200+ (0.106)	-0.089 (0.120)	-0.175 (0.117)
Observations	974	974	967	967
Adjusted R^2	0.054	0.207	0.029	0.158
Control variables		X		X

+ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$.

Table OA.2: Subsample analyses excluding missing: challenge-seeking

	Challenge		Difficult task		Advanced math ¹	
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: vocational track by GPA						
Treatment	0.115 (0.110)	0.119 (0.106)	0.023 (0.059)	0.026 (0.056)		
High × treatment	-0.056 (0.176)	-0.046 (0.173)	0.042 (0.081)	0.053 (0.078)	N/A	N/A
Observations	714	714	727	727		
Adjusted R^2	0.067	0.057	0.067	0.123		
Panel B: academic track by GPA						
Treatment	0.038 (0.102)	0.022 (0.107)	-0.007 (0.051)	-0.018 (0.051)	0.064 (0.061)	0.058 (0.046)
High × treatment	0.000 (0.145)	0.027 (0.156)	0.075 (0.067)	0.083 (0.064)	0.003 (0.064)	0.002 (0.057)
Observations	907	907	954	954	977	977
Adjusted R^2	0.002	0.044	0.065	0.170	0.243	0.405
Control variables		X		X		X

+ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$.

Table OA.3: Subsample analyses excluding missing: achievement

	Math quiz		GPA		Completed	
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: vocational track by GPA						
Treatment	-0.081 (0.087)	-0.071 (0.085)	0.028 (0.125)	-0.045 (0.117)	0.018 (0.032)	0.013 (0.031)
High × treatment	0.100 (0.149)	0.094 (0.132)	0.060 (0.156)	0.117 (0.135)	-0.030 (0.041)	-0.023 (0.039)
Observations	723	723	737	737	737	737
Adjusted R^2	0.104	0.230	0.395	0.537	0.087	0.154
Panel B: academic track by GPA						
Treatment	0.114 (0.128)	0.118 (0.115)	-0.066 (0.096)	-0.025 (0.069)	-0.049 (0.034)	-0.042 (0.029)
High × treatment	-0.113 (0.155)	-0.163 (0.141)	0.197+ (0.115)	0.119 (0.092)	0.083* (0.034)	0.071* (0.029)
Observations	939	939	981	981	981	981
Adjusted R^2	0.217	0.377	0.402	0.571	0.066	0.156
Control variables		X		X		X

+ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$.

Online Appendix B Cross tabulations

Table OB.1: Cross tabulation of pre-intervention mindset and track in high school. Survey 1 sample

	Fixed mindset	Growth mindset	Total
Vocational track	583	440	1 023
Academic track	578	755	1 333
Total	1 161	1 195	2 356

Note: n=2 364. 8 are missing due to not having answered the mindset questions.

Table OB.2: Cross tabulation of pre-intervention mindset and GPA. Survey 1 sample

	Fixed mindset	Growth mindset	Total
Low GPA	678	483	1 161
High GPA	443	675	1 118
Total	1 121	1 158	2 279

Note: n=2 364. 8 are missing due to not having answered the mindset questions and 77 are missing due to missing GPA in 10th grade.

Online Appendix C Distribution plots

Figure OC.1: Distribution plot: mindset measures. Survey 2 - Full sample (n=1 768)

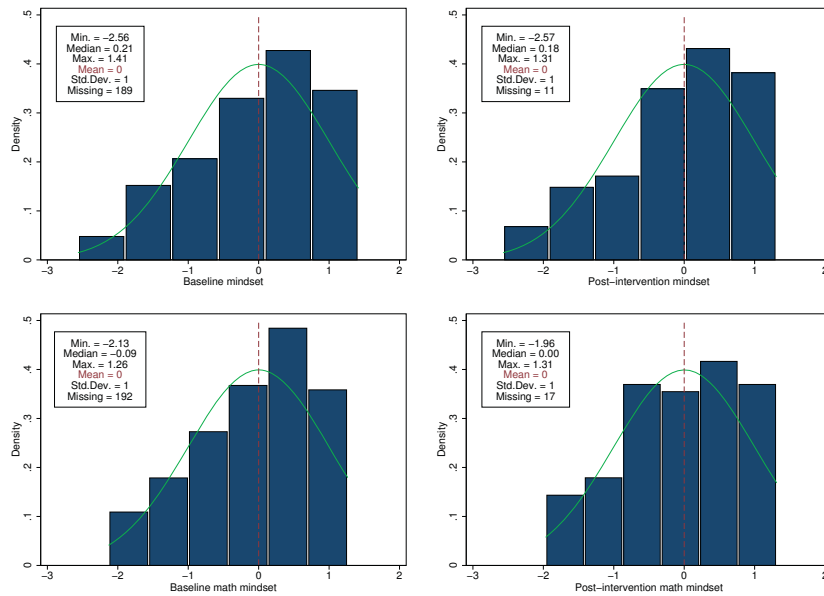


Figure OC.2: Distribution plot: mindset measures. Survey 2 - Vocational track sample (n=769)

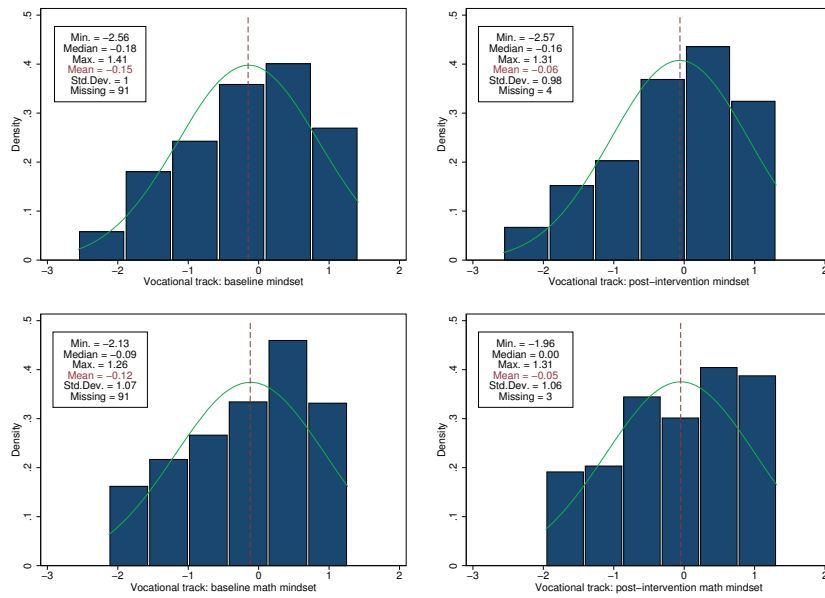


Figure OC.3: Distribution plot: mindset measures. Survey 2 - Academic track sample (n=999)

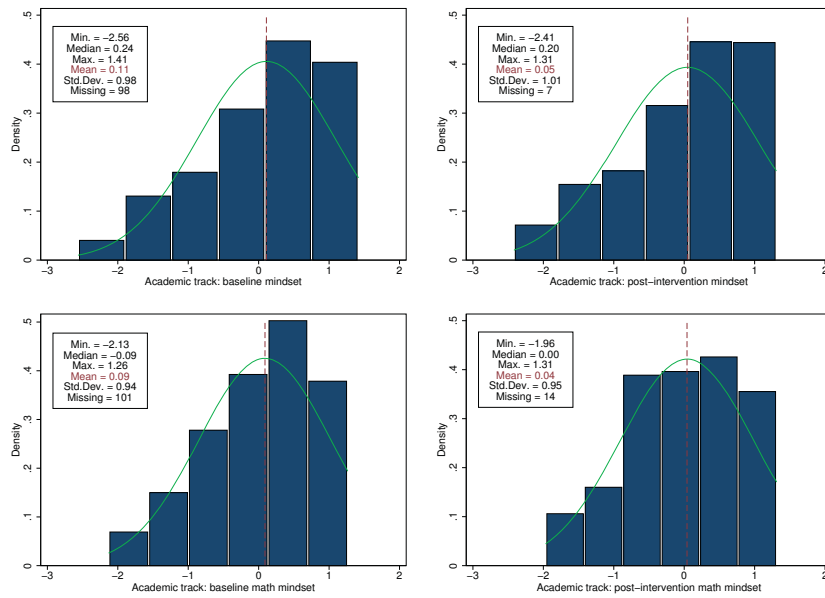


Figure OC.4: Distribution plot: baseline mindset measures. Survey 2 Sample (n=1 768)

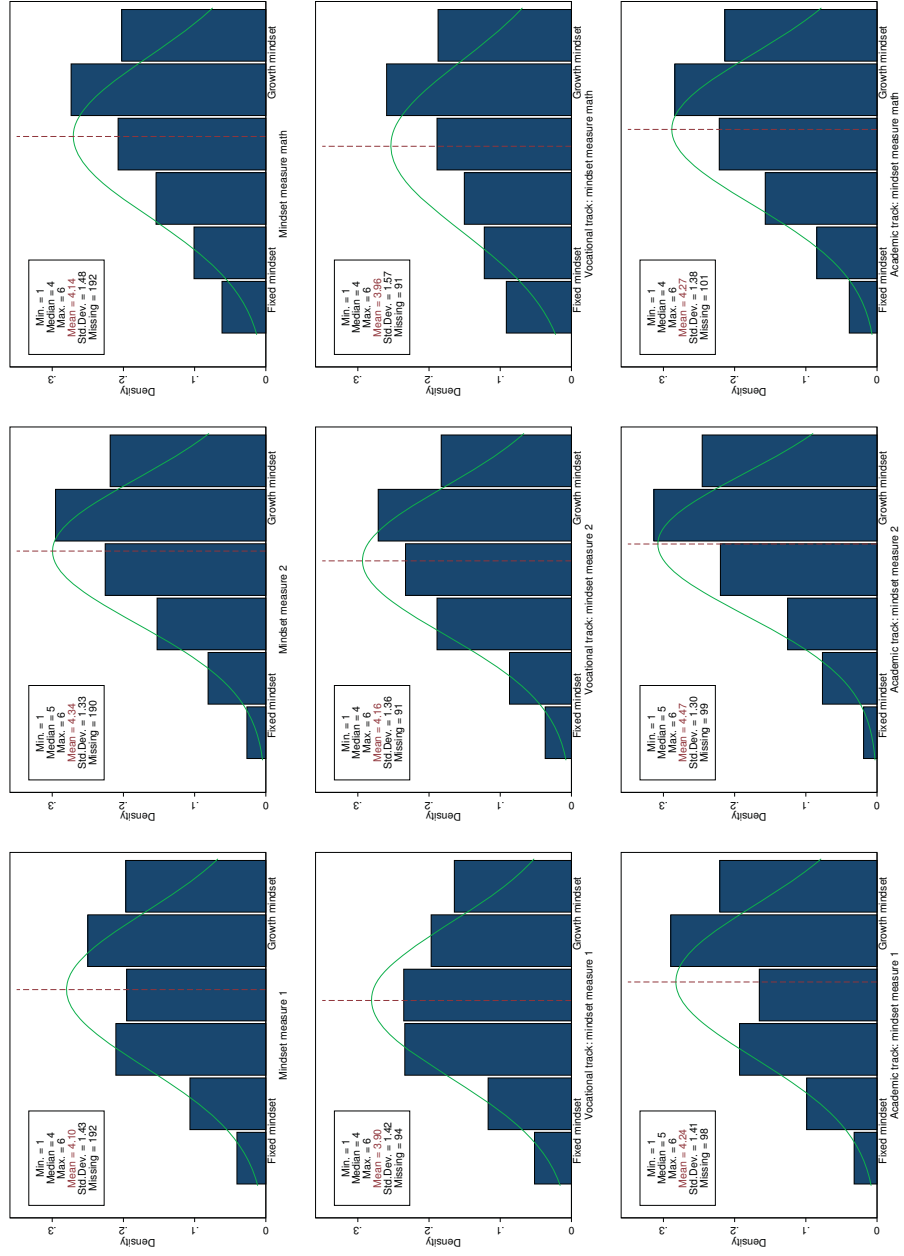


Figure OC.5: Distribution plot: post-intervention mindset measures. Survey 2 Sample (n=1 768)

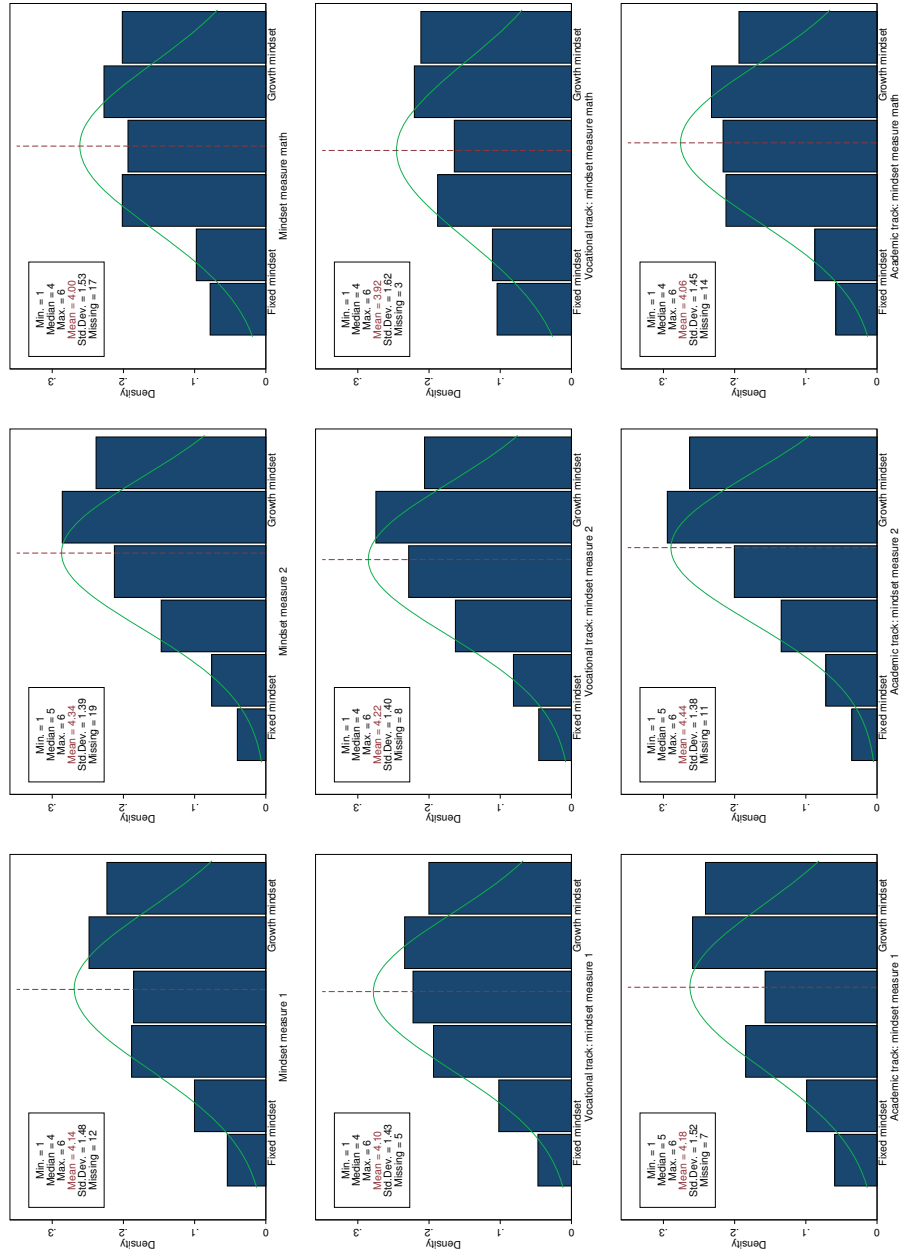


Figure OC.6: Distribution plot: challenge-seeking - "make-your-own-worksheet" task. Survey 2 Sample (n=1 768)

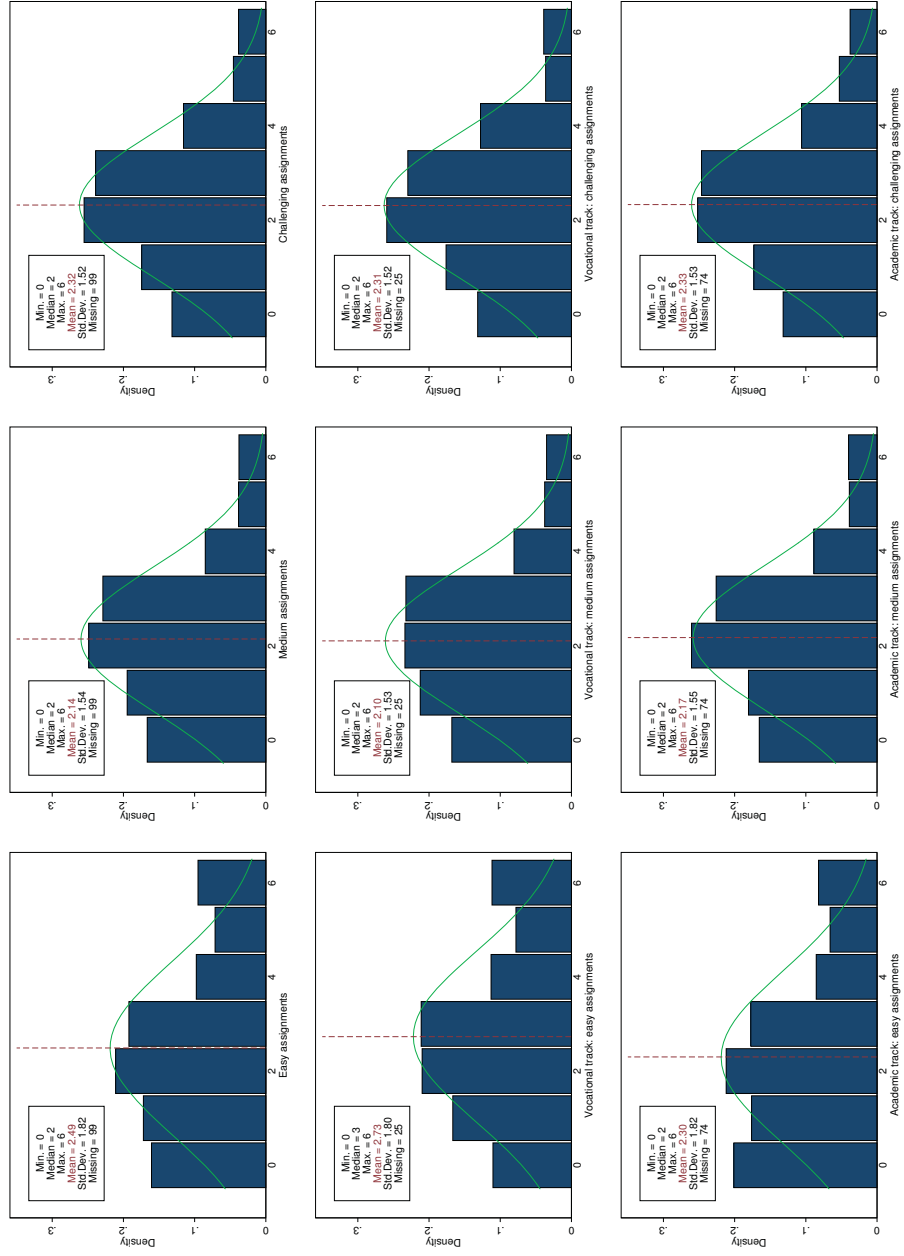


Figure OC.7: Distribution plot: achievement measure - math quiz. Survey 2 Sample (n=1 768)

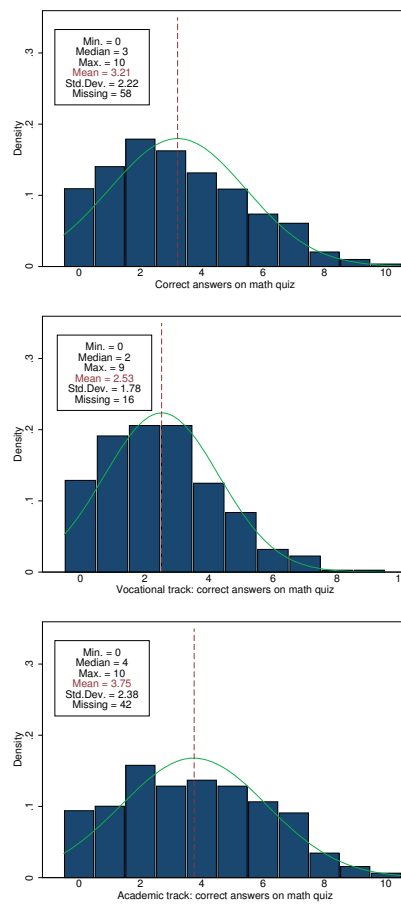


Figure OC.8: Distribution plot: achievement measure - grades. Survey 2 Sample (n=1 768)

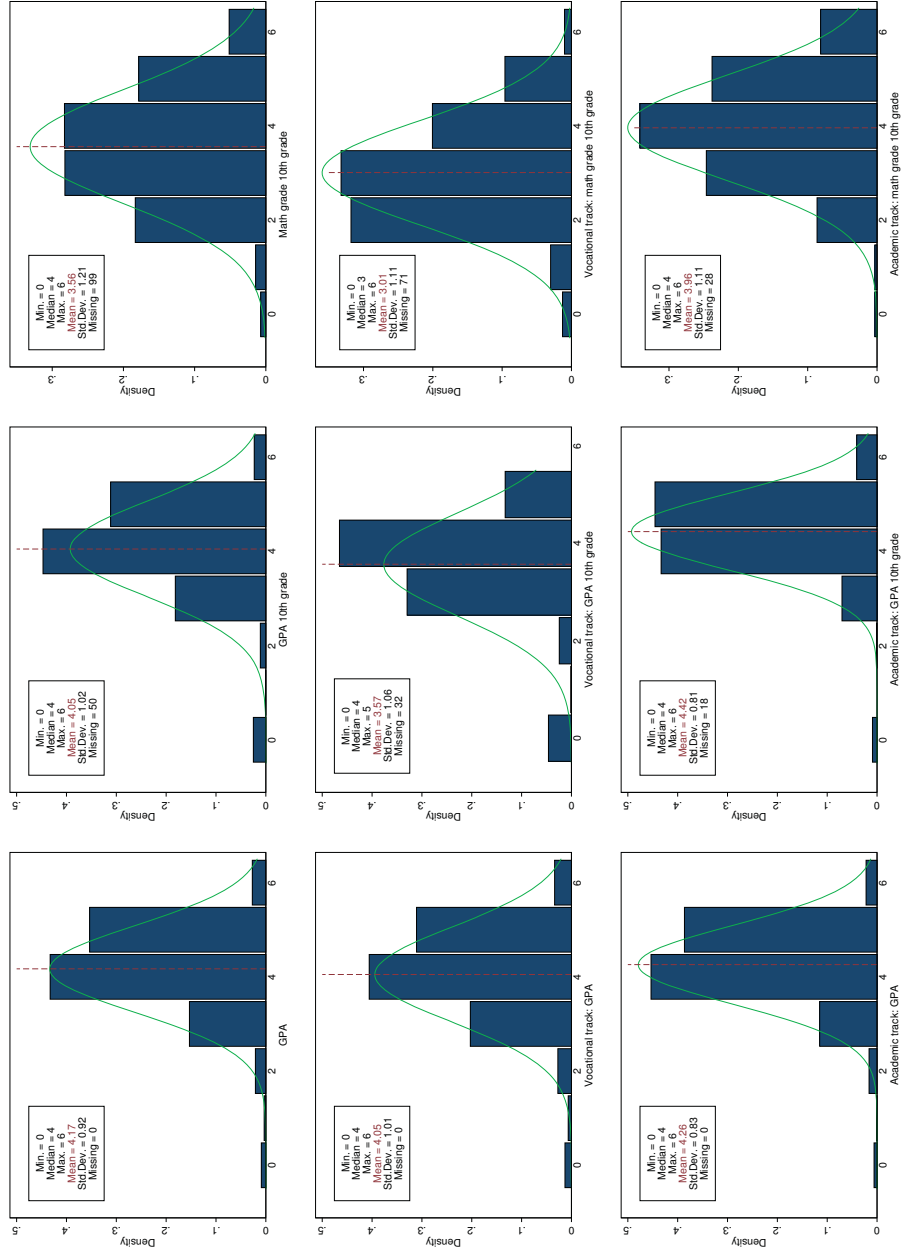


Figure OC.9: Distribution plot: attitudes and expectation measures survey 2. Survey 2 - Full sample (n=1 768)

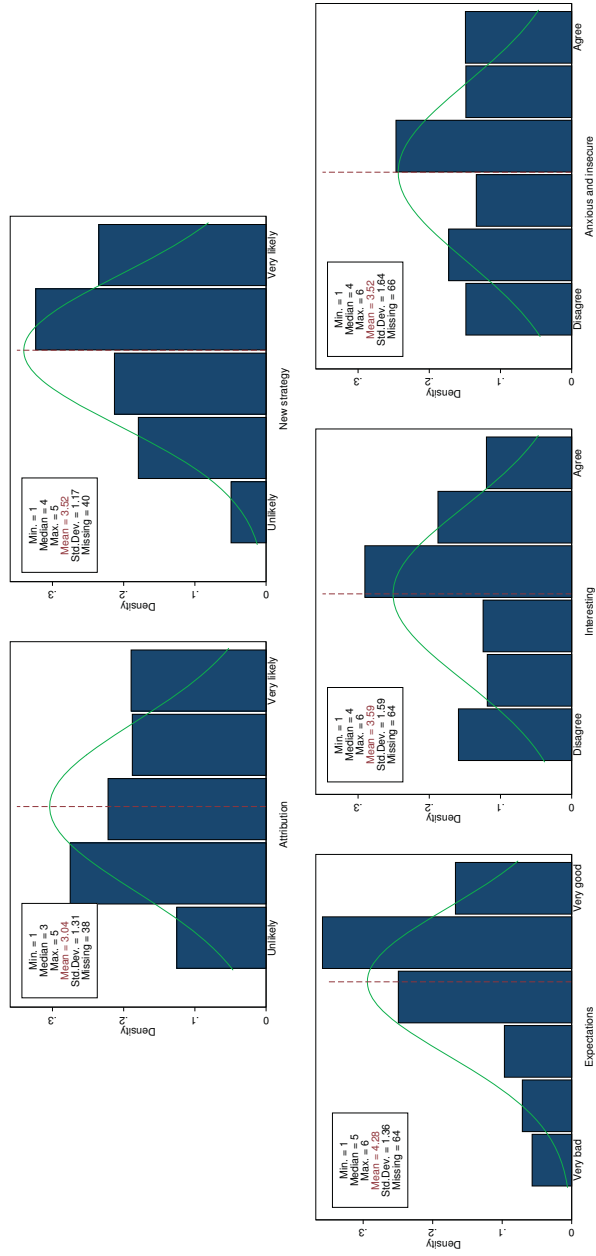


Figure OC.10: Distribution plot: attitudes and expectation measures survey 2. Survey 2 - Vocational track sample (n=769)

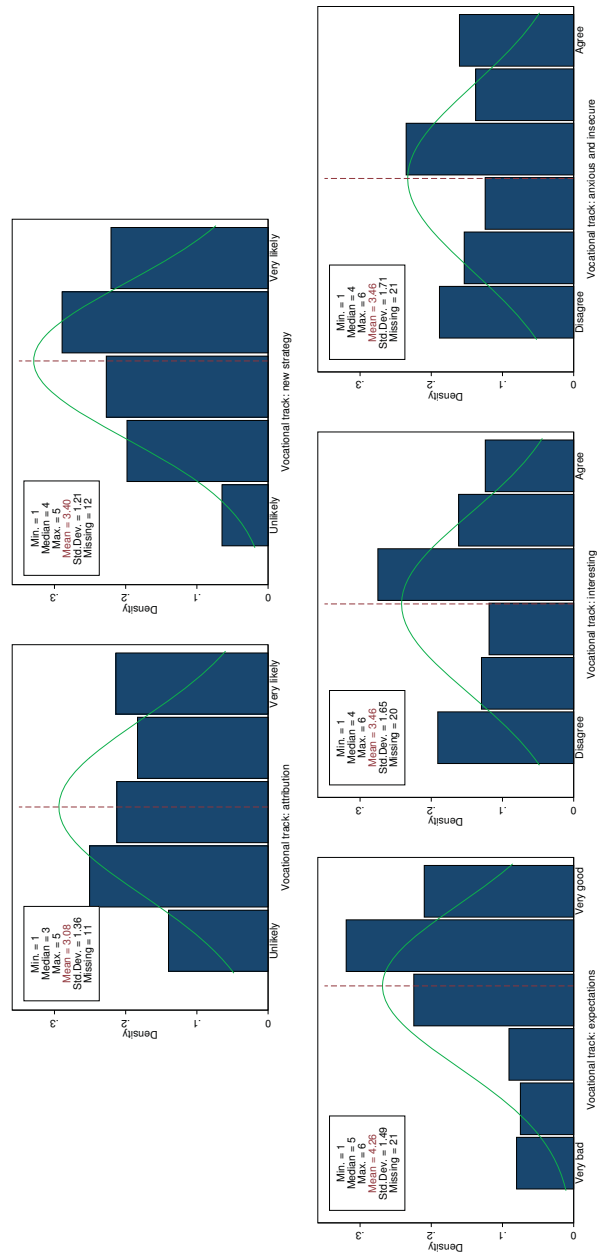


Figure OC.11: Distribution plot: attitudes and expectation measures survey 2. Survey 2 - Academic track sample (n=999)

