From K through 12 (or 13?)
Students' Health, Well-being, and Performance from Preschool to Graduation

Inaugural-Dissertation
zur Erlangung der Doktorwürde
— Doctor rerum politicarum —
an der Otto-Friedrich-Universität Bamberg

vorgelegt von

Johanna Sophie Quis, M. Sc.

2019

URN: urn:nbn:de:bvb:473-opus4-551598
DOI: https://doi.org/10.20378/irbo-55159

Kumulative Dissertation

Erstgutachter:	Prof. Dr. Guido Heineck
Zweitgutachterin:	Prof. Dr. Silke Anger
Drittgutachter:	Prof. Dr. Guido Schwerdt
Tag der mündlichen Prüfung:	14.05.2019
Dissertationsort:	Bamberg

Acknowledgments

Before embarking on this journey I was warned about the hard work and countless frustrations which result from doing a doctorate. Had I anticipated the true extent of agony and despair ahead of me, I would probably have quit. Not knowing how bad it would be, I held it with Dory and "just keep swimming" became my mantra. Whenever that did not work, I was certain to have someone supporting my back and lifting me back up. Without the continuous support of many of you, I would therefore not be in a position to finally finish off the dissertation with these opening words.

First and foremost I would like to thank my first supervisor Guido Heineck for giving me the chance to pursue my dissertation in Bamberg. His ongoing support, incredible ability to see people, appreciation, and advice in professional and personal questions guided me through the past years. I am also grateful to Silke Anger who agreed to be my second supervisor and provided helpful comments on my research. Finally, I would like to thank Guido Schwerdt, who agreed to be my third supervisor. Moreover, I am indebted to my co-authors Stefanie Herber, Simon Reif, and Anika Bela.

I would also like to thank the Bamberg Graduate School of Social Sciences (BAGSS) for funding me and providing a fruitful research environment. Within BAGSS I am especially indebted to Miriam Schneider and Marc Scheibner, who relentlessly supported us BAGSS students and took genuine interest in our well-being and professional development.

Uta and Saikat started this crazy journey with me, were great office mates and good to talk to when life happened while we were making other plans. Angelika was always a role model for being a mother and doctoral student at the same time and also proved to be an amazing office mate. Our BAGSS lunch group (Gundula, Irina, Martin, Stefanie) regularly lifted my mood and listened to me ranting about life's hardships. Stefanie, next to being an excellent travel mate, became a special friend in many regards. I admire Susanne Elsas for her calm spirit and I would not have wanted to miss the year we spent together at the chair. Both in and outside Bamberg, I am lucky to have and have found friends (Jana, Jasmin, Laura, Barbara and Andreas, Bianca and Nico, Dorothea and Oliver) who were always there.

Finally, I want to thank my family. My parents were very patient with my educational career and supported me in all dimensions. I always love coming home to see my siblings. Special thanks go to my own little family. Sönke is way too small to understand the implications of pursuing a doctorate, even though, of course, he is already noticing my being at work quite a lot. But whenever I come home or pick him up at daycare his complete focus on the basic things provides me with a lot of joy and happiness. Helge, more than anyone else, bore with the implications of my endeavor and lovingly supported me without conditions. I appreciate your endless levels of patience and understanding, especially in this final phase of this dissertation. Thank you.

Bamberg, August 13, 2018
Johanna Sophie Quis

Contents

List of Figures

List of Tables

List of Abbreviations

ATE Average treatment effect
BMI Body mass index
DiD Difference-in-differences
DST Daylight saving time
G8 German academic track duration of eight years
G9 German academic track duration of nine years
HLM Hierarchical linear models
IEA International Association for the Evaluation of Educational Achievement
NEPS National Educational Panel Study
OECD Organisation for Economic Co-operation and Development
OLS Ordinary least squares
PIRLS Progress in International Reading Literacy Study
PISA Program for International Student Assessment
RCT Randomized controlled trial
RDD Regression discontinuity design
SAT Scholastic Aptitude Test
SOEP German Socio-economic Panel
ST Standard time
TIMSS Trends in International Mathematics and Science Study

Chapter 1

Introduction

Education is seen as a critical determinant of individual and societal flourishing and is hence constantly on policymakers' agendas. For example, the Europe 2020 strategy includes eight headline targets, two of which are related to educational attainment, reducing school dropout and increasing the share of university graduates.

Further, one main topic of discussions about education in Germany and many other OECD countries concerns the capabilities of educational systems to raise student achievements. The self-identified nation of poets and thinkers was shocked by the mediocre results from the first round of the Program for International Student Assessment (PISA) in 2000. As a result, education became a major topic of German policy and public discussion. At the same time, demographic change and a relatively low female labor market participation also affected educational policies. In the past two decades, German educational policy changed in many dimensions: The right to early childhood education was established and some federal states abandoned kindergarten fees. Disabled children were more frequently placed in regular kindergartens and schools, tracking mechanisms between primary and secondary school changed and structural changes in lower track secondary schools were widely implemented. The academic track secondary school was compressed by a year. However, this compression was largely unpopular and most West German states returned to the old system or a more flexible version thereof. Finally, the bologna reforms and excellence initiative changed tertiary and postgraduate education. Thus, educational policy has changed recently to a certain degree at all stages of educational careers yielding a large scope for policy evaluation.

This dissertation in the economics of education consists of four independent research articles, which are oriented along the life course of students' school careers. The first study assesses the relationship between non-cognitive skills and competence in kindergarten and early primary school. The three following studies assess different policies' impacts on student outcomes: First, the effect of a shock to sleep duration on primary school student performance in international large scale assessments is evaluated, followed by the relationship between the academic track compression and students' physical and mental health at academic track graduation and the following years.

The idea of education being a valuable good dates back a long time. However, it has been coined for modern economics about sixty years ago. Mincer (1958, 1974) initiated the empirical estimation of monetary returns to education, Schultz (1959, 1960, 1961, 1962) and, more formalized, Becker (1962, 1993) introduced *human capital* which, equivalent to physical or financial capital, is formed through investments into an initial stock. Typically education is seen as the most important investment good into human capital. For example, Becker (1993) states education is reflected in wage differentials, since the latter reflect productivity differentials, which in turn are determined by differentials in educational attainment. Since then, the modeling of human capital has changed. Most notably, Ben-Porath (1967) added a lifetime perspective, according to which people can choose when and for how long they invest into their education and extended the model to the possibility of training, i. e. education after formal education is finished. Cunha and Heckman (2007, 2008) formalized a theoretical model of the process of skill formation.

In contrast to the perspective that education increases human capital, signaling theory (Spence, 1973) rejects the idea of education improving human capital. Instead, individuals are endowed with initial skill levels and educational attainment merely reflects other characteristics of the individual for which she may be rewarded on the labor market. If this assumption were true, most aims of educational policy (like reducing inequalities and improving skills) would be unattainable. Even though there is some evidence of *sheepskin effects*, i. e. degrees are honored more than attaining the same amount of education without the final certificate (Hungerford and Solon, 1987; Jaeger and Page, 1996), there are also several studies that claim a causal relationship between education and several dimensions of human capital.

Education is linked to a multitude of beneficial outcomes at the societal and individual level. For example, education facilitates innovation and productivity of the society if a more educated population is better capable of adopting new technologies and thereby increasing growth. On the country level, nations with more skilled populations perform better in terms of economic growth (Barro, 2001; Hanushek and Wößmann, 2008, 2012; Balart et al., 2018). On the individual level, educational attainment is linked to lower crime rates (Lochner and Moretti, 2004; Machin et al., 2011), more political interest (Dee, 2004; Milligan et al., 2004; Siedler, 2010), better health (Brunello et al., 2013; Silles, 2009) and health behaviors (Currie and Moretti, 2003; de Walque, 2007; Cutler and Lleras-Muney, 2010; de Walque, 2010), and better labor market outcomes (Card, 1999). These findings strongly support the beneficial effects of education.

On both the individual and the country level, education is always a tradeoff between costs that occur in the present and benefits that occur in the future. On the country level, direct costs usually include the provision of education. Indirect costs occur due to foregone income tax and lower contributions to the social security systems as well as the opportunity of the individual's skills not being employed elsewhere to produce positive externalities.[1] The delayed benefits are potentially higher income taxes in the future, lower health care costs, and so on. On the individual level the direct costs encompass possible monetary costs (tuition fees, books, other education-related expenditures) and mental costs (lower utility, learning is hard work). Indirect costs are for example foregone earnings from not working while attaining further education. The benefits in the long term, among others, stem from higher earnings and better health.

Individuals tend to be bad at intertemporal decision making and, in addition, do not calculate the societal benefits of their own education. Hence, they are likely to invest too little into their own educational attainment, which serves as a common argument for governments to finance or subsidize education (Hanushek, 2002). It is important to balance tailored education to let students reach their full potential and keeping education affordable to individuals and to the society.

Even in the beginnings of human capital theory, researchers were aware that human capital encompasses a broad range of capabilities beyond formal education. Schultz proclaimed a definition of human capital "[...] represented by training, education, [and] additional capabilities based on health and new knowledge" (1959, p. 114). Similarly, Becker (1962) already was aware that investments into human capital, which serve to improve "the physical and mental abilities of people and thereby raise real income prospects" (p. 9) include more than just education. Human capital rather needs to be interpreted as a broad construct that includes all characteristics that enable individuals to generate economic values. This includes cognitive and non-cognitive

[1] The lack of positive externalities might be compensated by positive externalities from educational institutions, like universities.

skills[2], personality, and health. Grossman (1972) proposed health as another important outcome of human capital formation through education.

Non-cognitive skills are important for individual success and manifest during childhood and adolescence. The development of competences usually takes place throughout childhood (e. g. Cunha and Heckman, 2008). However, there is evidence that non-cognitive skills are malleable for a longer period than cognitive skills (Brunello and Schlotter, 2011 provide an overview of the relevant literature). In their model of skill formation, Cunha and Heckman (2007) propose pathways through which levels of and investments into skills (cognitive and non-cognitive) affect skills in the following periods and thereby human capital accumulation. In this setting skill begets skill, i. e. higher initial levels of skill foster larger productivity of investments into these skills. Additionally, through a multiplier process, the level of one skill also affects the level and growth of other skills. In this framework, early investments are particularly fruitful and increasing non-cognitive skills might also increase cognitive performance. Non-cognitive skills are often of better explanatory value for lifetime outcomes than cognitive skills (e. g. for the Perry Preschool program: Heckman et al., 2013), and often are used to predict educational attainment and other lifetime outcomes (e. g. Mischel et al., 1989; Duckworth and Seligman, 2005; Heineck and Anger, 2010). Hampson et al. (2007) show that childhood personality traits affect adult health status directly and indirectly through different eating habits, educational attainment and smoking behaviors. Balart et al. (2018) also show that non-cognitive skills of a population affect economic growth to a similar degree as cognitive skills do.

The present dissertation, next to cognitive skills represented by competence measures, emphasizes non-cognitive skills and health. It employs micro-econometric techniques to assess aspects of the relationship between non-cognitive skills or institutional settings and competences in the first part, and the relationship between a change in institutional settings and students' mental and physical health in the second part.

The fundamental interest in empirical research is to identify causal effects of an input on an outcome. While chapter 2 is generally rather descriptive, I claim to capture causal effects in chapters 3, 4, and 5. It would be ideal to observe the same individuals with and without the respective treatment, which generally is not possible. Therefore, special care needs to be taken when trying to identify causal effects and claiming causality (Angrist and Pischke, 2009). The second-best approach, a true randomized controlled trial (RCT) is also close to impossible in the evaluation of educational policies, mostly for financial and ethical reasons.[3] Therefore, I rely on the allocation of students into treatment and control groups dependent on factors they cannot control, e. g. date of birth, in order to derive causal claims.

In the following paragraphs, I outline the contents and contributions of the chapters included in the thesis (see also table 1.1). Further, I discuss implications of the findings and possible limitations.

The first part of this dissertation (chapters 2 and 3) is concerned with modeling determinants of educational achievements via an educational production function. The educational production function incorporates student characteristics and family background, as well as school factors and institutional settings as determinants of students test achievements. Student and family characteristics typically encompass cognitive and non-cognitive skills, basic demographics, and socio-economic variables of parents and children. In the past, institutional factors that have received a lot of attention were school autonomy (Clark, 2009), compulsory school dura-

[2] Strictly speaking, the division of skills into cognitive skills (intelligence, subject specific competence) and non-cognitive skills (social skills, self-regulation, ...) is not correct as there are only very few non-cognitive skills that indeed do not induce cognitive processes (Duckworth and Yeager, 2015) skills, but the terms are generally used in the literature and I will stick to them.

[3] Although there are some examples mainly from the US where RCTs were conducted (e. g. STAR, Perry Preschool Project).

Table 1.1: Overview of dissertation

	Chapter 2	**Chapter 3**	**Chapter 4**	**Chapter 5**
Title	Preschoolers' self-regulation, skill differentials, and early educational outcomes	Does the Transition into Daylight Saving Time Affect Students' Performance?	Does higher learning intensity affect student well-being? Evidence from the National Educational Panel Study	Health Effects of Instruction Intensity: Evidence from a Natural Experiment in German High-Schools
Data	NEPS SC 2	TIMSS, PIRLS	NEPS BW	SOEP
Methods	OLS	Hierarchical linear models applying Rubin's rules for imputed data in a setting closely mimicking the mechanics of a regression discontinuity design	Before-after comparison in a quasi-natural experiment	Difference-in-differences estimation, triple difference-in-differences estimation
Co-author(s)	Anika Bela and Guido Heineck	Stefanie P. Herber and Guido Heineck	–	Simon Reif
Own contribution	45%	47.5%	100%	50%

tion (Angrist and Krueger, 1991; Brunello et al., 2009; Clark and Royer, 2013), and the degree and timing of tracking (Hanushek and Wößmann, 2006; Bauer and Riphahn, 2006; Brunello and Checchi, 2007). At the school level, research mainly concerns class size (Hanushek, 1999, 2006) and teacher characteristics (Rivkin et al., 2005; Hanushek and Rivkin, 2006, 2010).

The first study in this dissertation is focused on a student characteristic as input into the educational production function. **Chapter 2** was jointly prepared with Anika Bela and Guido Heineck. We analyze the relationship between preschool children's self-regulatory skills, i. e. patience, and their mathematical competence and competence development over early primary school. The chapter uses panel data from the kindergarten-cohort of the National Educational Panel Study (NEPS). It builds upon and extends the work by Lorenz et al. (2016) by adding a longitudinal perspective. While we find significant differences in the initial levels of mathematical competence between patient and impatient children, there seems to be no effect of patience on the development of competences in the full sample. Our results are robust to the inclusion of measures for general cognitive skills. However, we are able to show that among those with lower initial mathematical competence, the patient children are able to lessen the gap relative to children with high initial math competences. To the best of our knowledge, we are the first to assess the longitudinal mechanics of the interplay between self-control and cognitive skills in early childhood for German children.

Education is often understood as a means to reduce inequality by increasing chances for children from disadvantaged backgrounds to acquire the same level of skills as children from more advantaged backgrounds. However, studies show that inequalities arise early and differences between children from disadvantaged backgrounds and children from more advantaged backgrounds are almost impossible to close by the time they enter school (e. g. Heckman, 2006). Our results point into a similar direction and are in line with findings by Heckman et al. (2013) who find that differences in outcomes of the Perry Preschool program are largely driven by non-cognitive skills. In order to reduce inequalities, interventions need to take place as early as possible during childhood and should ideally target both cognitive and non-cognitive skills.

However, evidence on types of interventions which could improve children's patience is very scarce. This lack of knowledge provides scope for further research which hopefully could inform policy.

While chapter 2 assessed the relationship between a student characteristic and competence in the educational production function, **chapter 3** assesses an institutional factor. School start times have received only limited attention in the literature (Carrell et al., 2011; Hinrichs, 2011; Edwards, 2012; Heissel and Norris, 2017). Stefanie Herber and Guido Heineck and I again assess student competence, this time focusing on possible unintended consequences of a policy. We are the first to examine the effect of the transition to daylight saving time (DST) in spring on student performance in international low stakes student assessments. The channel through which we expect the transition to DST to affect student competences is via sleep deprivation due to the hour that is lost in the night between Saturday and Sunday. We utilize a quirk in the *Trends in International Mathematics and Science Study* (TIMSS) and the *Progress in International Reading Literacy Study* (PIRLS) tests in 2011. Both studies were conducted in the weeks surrounding the switch to DST in six European countries, providing data from more than 22,000 students. Using an estimation technique loosely mimicking a regression discontinuity design (RD), we compare the test performance of students who were randomly allocated to test dates before or after the switch to DST. In this setting we account for potential non-linearity in test performance (e. g. due to a blue Monday effect). Due to the hierarchical structure of the data and since the outcome measures are provided as several plausible values for the competence of each individual student, we estimated the effects with hierarchical linear models (HLM) that additionally account for uncertainty in the imputed outcomes applying Rubin's rules. The results suggest a DST-effect, which is small in magnitude and not statistically significantly different from zero. A broad range of robustness tests, for example classical RD estimation, serve to confirm the zero-effect. We therefore conclude that the switch to DST does not harm student performance and therefore does not challenge the validity of conclusions drawn from student assessment data collected around the switch to DST.

In conclusion, the first part of this dissertation was considering different inputs into the educational production function. The second part shifts the perspective away from the classical educational production function towards the broad view of human capital including health and well-being. Chapters 4 and 5 evaluate a series of educational reforms in Germany. The G8 reforms implied a reduction in school years needed to attain a degree from the academic track from nine to eight years of secondary school, while the federal requirements of cumulative instruction hours and contents to be taught remained unchanged. The scientific evaluation of these reforms focused mainly on student's grades (Büttner and Thomsen, 2015; Huebener and Marcus, 2017), cognitive (Dahmann, in press; Huebener et al., in press; Hübner et al., 2017) and non-cognitive (Thiel et al., 2014; Dahmann and Anger, 2014) skills, educational decisions (Meyer and Thomsen, 2016; Meyer et al., forthcoming; Marcus and Zambre, forthcoming) and performance in university (Dörsam and Lauber, 2015; Meyer and Thomsen, 2017). The public discussion of the reforms were focused on the extent of burdens students were facing due to the reforms. An early study on burdens of the G8 system was comparing different systems across states (Böhm-Kasper and Weishaupt, 2002), in this setting causality is hard to establish, as the differences might also be attributed to inter-state differences. A more recent branch of the literature claims to assess causal effects on burdens, health and health behaviors (Meyer and Thomsen, 2015; Westermaier, 2016; Meyer and Thomsen, 2017; Hofmann and Mühlenweg, 2017). I add to this literature by shifting the perspective to stress and internalizing mental health problems as well as subjective well-being.

In **chapter 4**, I assess students at the end of their school career close to high school graduation. The reform outcomes evaluated include students' perceived stress, internalizing mental health problems and well-being. The study builds upon data on over 2,300 students from the National Educational Panel Study's (NEPS) Additional Study Baden-Wurttemberg. The data was collected in the German federal state of Baden-Wurttemberg with the purpose of evaluating the impacts of the G8 reforms in Germany. I argue that the reform serves as natural experiment, where the assignment to the treatment or control group is random conditional on the year of school entry, which took place long before the reform was announced. The estimation of the effects relies on ordinary least squares (OLS). The results imply an increase in perceived stress and symptoms of internalizing mental health problems for females. For males, I only find an increase in perceived stress. However, neither females nor males are affected by the reform in terms of more general well-being, as measured by well-being in school and life in general. The robustness checks include several different ways of constructing the indexes that form the dependent variables and using the last pure G9 cohort and the first pure G8 cohort instead of the double cohort for my analyses.

While the data provides a relatively large sample size, excellent measures for perceived stress and internalizing mental health problems and a large set of controls, the nature of the reform in Baden-Wurttemberg only allows to compare different cohorts, which are either affected by the reform or not. It is therefore impossible to separate general time trends from the reform effect. This would be a problem if any other factor changed between the cohorts and affected the treated and untreated students in different ways. This problem is addressed in **chapter 5**, jointly prepared with Simon Reif. We assess the G8-reforms in most German states focusing on physical and mental health outcomes. We use data from the German Socioeconomic Panel (SOEP) and examine the effects of the reform on students aged about 17, who are still in school at the time of the interviews and students who already graduated to see whether the health effects of the reforms are transitory shocks or persist after students finished school. The sequential introduction of the reform over several years in the different states in Germany allows for a difference-in-differences (DiD) design of our study. In contrast to chapter 4, we are therefore able to rule out that our results are pure cohort effects. We use a triple-DiD approach with students from the intermediate secondary school track as additional control group. Further, we provide extensive robustness checks analyzing effect heterogeneity, bootstrapped standard errors (Cameron et al., 2011), and a special kind of placebo test, the permutation test suggested by Chetty et al. (2009). The results again imply significant effects of the reforms on female student's body mass index (BMI) and worrying while they are still in school. We find no effect on any health measure for male students. In the sample of graduates we do not keep age constant, but instead the distance to graduation. The observed individuals graduated either one or two years before the interview, dependent on data availability which is not related to the assignment to the treatment or the control group. In the analyses of graduates, we are able to use more thorough health data, which is collected biannually from the adult SOEP sample. We find no impact on mental and physical health for male graduates and in some specifications positive effects on females. The positive effects on females after graduation are in line with findings on university students by Meyer and Thomsen (2017). However, we cannot confirm their finding of worse health for males who attended G8. While we were preparing this chapter, Hofmann and Mühlenweg (2017) also using the SOEP evaluated a pooled sample of students and graduates finding a slight decrease in mental health but no effect on physical health or smoking behavior.

The G8 reforms only date back a few years, and therefore, it is not yet possible to evaluate true long-term outcomes of the reforms. The short term results seem to allow the interpretation that the increase in instruction intensity might affect students while they are still in school but the effects level off soon after graduation. Additionally the effects seem to stem mostly from students in the transition phase where the reform implementation might have been rough, before teachers had fully adjusted their lessons to new settings. Finally, the double cohorts faced increased competition due their larger size. As a result, it would be surprising to find long term health effects of the reforms.

While research on G8 finally has been established and most findings are suggesting that the system matters less than whether or not a student is affected by the transition, most West German states are already transitioning back to nine-year-systems. These new systems usually offer more flexibility to choose between finishing in eight or nine years than before the original G8 reforms. The move back to nine-year systems will result in gap years with either no or very small graduation cohorts in the respective states – another interesting topic for further research in the German educational system.

Chapter 2

Preschoolers’ self-regulation, skill differentials, and early educational outcomes

Johanna Sophie Quis, Anika Bela and Guido Heineck

2.1 Introduction

Individuals’ capabilities of delaying or even foregoing immediate consumption in order to yield better future outcomes is a critical behavioral component in life. Underlying mechanisms and processes, however, are differently addressed across economics and psychology. In economics, the “rate of time preference” is the best known concept to reflect individuals’ degree of patience and is one of the most relevant theoretical parameters for modeling future-oriented, intertemporal processes, including investment decisions, savings, health behavior, or human capital accumulation.

The psychological literature too has a long history of interest into individuals’ underlying self-regulatory skills (Vohs and Baumeister, 2016), how they relate to observable heterogeneity in, for example, delaying gratifications and whether they predict different life outcomes. A particularly well established literature investigates how children differ in self-regulation and how these differentials explain e. g. adolescents’ or adults’ social and cognitive outcomes: Since the late 1960s, analyses based on the now famous *Marshmallow test*[1] (Mischel et al., 1989) and numerous follow-up studies suggest for higher performance and better outcomes of individuals, who in their childhood were more patient, through their mid-forties (Casey et al., 2011; Moffitt et al., 2011).[2]

Another line of research in economics addresses the elicitation of adults’ time preferences (e. g. Frederick et al., 2002; Andersen et al., 2008) or individuals’ health and health-related behavior (Courtemanche et al., 2015; Bradford, 2010). Little is, however, known about the time preferences and their impact for teenagers and, even sparser, for children. Spurred by the work of Heckman (e. g. Cunha and Heckman, 2007), recent research started to explore whether early differentials exist and by how much they affect (or are at least correlated to) later-life outcomes (Golsteyn et al., 2014).

We contribute to this yet scarce research by examining the relationship between preschoolers’ delay of gratification, which is a manifestation of individuals’ self-regulation (Neubauer et al.,

[1] With an interest in the underlying psychological processes, the test aimed at assessing children’s ability to delay gratification for a bigger reward (Mischel and Ebbesen, 1970; Mischel et al., 1972, 1989): Children were seated at a table and they were offered a marshmallow (or a similar food item that the child liked) that was set directly in front of them. The tester instructed the child that it could either wait until the tester returned and get a second Marshmallow or eat the one available before the tester returned but would in this case not get another one. The recorded waiting time was then interpreted as measure for children’s self-imposed delay of gratification.

[2] In a replication study, Watts et al. (2018), however, challenge this narrative. As will be outlined in the section on prior research, they find smaller effects of children’s waiting times on behavioral outcomes at age 15 which even vanish as soon as additional background factors are accounted for.

2011) and their mathematical competence and its development early in primary school. Addressing children's mathematical competences is relevant because early math skills are a major determinant, if not a causal factor, for adolescents' school success (e. g. Watts et al., 2014, 2017), which then, on average, contributes to better adult life outcomes.

Another contribution is the use of data from the kindergarten cohort of the German National Educational Panel Study (NEPS)[3], which is a recent and rich data source on individuals' competences and their competence development. So far, barely any research has explored NEPS for the questions addressed here. Using this data allows us both to account for a broad set of relevant background variables for a sample of children from diverse social backgrounds. We further extend the study of Lorenz et al. (2016), who look at the impatience-skills-nexus from a cross-sectional and thus more descriptive perspective, by using the longitudinal dimension of the data.

We find a significantly positive association between children's self-regulation and their mathematical competence levels, even when holding general cognitive ability in kindergarten constant. Self-regulation is however not related to competence development over the first two years of primary school, meaning that the initial skill gap neither widens nor narrows substantially. Heterogeneity analyses imply that self-regulation benefits children with low initial levels of mathematical competence at the transition from kindergarten to primary school. This advantage, however, vanishes between grade 1 and grade 2.

2.2 Background and Prior Research

2.2.1 Background

Conceptually, our study aligns with elements from the model of skill formation by Cunha and Heckman (2007). According to this model, variation in skills is the result of *self-productivity* and *dynamic complementarity*, meaning that the stock of skills at a particular stage in life is a function of all past investments: While self-productivity implies that past skills increase later skills directly, dynamic complementarity increases the productivity of investments into skills for individuals with a higher prior level of skills.

The data we use do not allow to assess the two mechanisms to full extent, mainly because they do not provide information on investment in the skills we are interested in. Beyond that, we focus on the effect of one skill on another, so that we examine what Cunha and Heckman (2007) call cross-effects: *Cross-productivity* displays the effect of the level of one type of skill in the initial period on the level of another skill in a future period while *dynamic cross-complementarity* suggests that investments into the other skill are more fruitful if the person had a higher level of the respective skill in the initial period.

In our analyses, we first assess the relationship between children's self-regulation/patience, for which we have only cross-sectional information, and the level of their mathematical competence as well as its short-term development. Based on the cross-productivity notion, we expect higher mathematical competence for more patient children.

We should also see that differences in initial skill levels explain different gains in mathematical competence over time, possibly driven by all four mechanisms: Self-productivity implies that

[3] This paper uses data from the National Educational Panel Study (NEPS): Starting Cohort 2 – Kindergarten (From Kindergarten to Elementary School), doi:10.5157/NEPS:SC2:5.1.0. From 2008 to 2013, NEPS data were collected as part of the Framework Programme for the Promotion of Empirical Educational Research funded by the German Federal Ministry of Education and Research (BMBF). As of 2014, the NEPS survey is carried out by the Leibniz Institute for Educational Trajectories (LIfBi) at the University of Bamberg in cooperation with a nationwide network.

higher initial math competences positively affect future math competences. Dynamic complementarity suggests that investments into mathematical skills yield higher competence increases for children who start from a higher competence level. In line with cross-productivity, children, who are more patient in kindergarten, attain higher mathematical competences, and dynamic cross-complementarity finally triggers that patient children profit more from investments into their mathematical skills. Again, as there is no information on investments and as we have only a cross-section measurement of the child's self-regulation, we cannot directly test the Cunha-Heckman model, but we rather think of it as conceptual guideline.

2.2.2 Prior Research

Research on individuals' self-regulatory skills, or patience, in both psychology and economics can be grouped by its respective interest, i. e. whether the studies examine determinants of self-regulation/patience, its use as predictor of life outcomes, or whether children's intertemporal choice behavior can be influenced.[4]

As for determinants, both nature and nurture play a role for how children differ in self-regulation in their first years of life.[5] Children's age, reflecting their brain development and its effects on decision processes, is a critical factor (Sutter et al., 2015; Bartling et al., 2010), as are children's birth weight, their cognitive skills (Bartling et al., 2010), or breastfeeding duration (Falk and Kosse, 2016). Family background matters as well: Bartling et al. (2010), for example, use data from the German Socio-economic Panel (SOEP) and refer to the importance of maternal patience, which may hint towards a genetic component in children's initial skill endowment, but they also refer to the importance of parental assets, including house-ownership or number of books at home. Exploring data from NEPS, Lorenz et al. (2016) also find that patience increases with age, and that girls are more patient than boys. With respect to socio-economic background they find that children with educated parents tend to be more patient and children with both parents born abroad are slightly more patient.

As for outcomes, there is by now abundant evidence that individuals' patience is related to a large set of socio-economic indicators. To start with, results from the initial marshmallow tests show that more patient children, i. e. preschool children who were able to delay gratification for more time, performed better on a variety of outcomes throughout adolescence and adulthood: More patient children had a lower body mass index (BMI) (Schlam et al., 2013; Seeyave et al., 2009), and performed better on a test of cognitive control during adolescence (Eigsti et al., 2006) and even in their mid-forties (Casey et al., 2011).

Recently, Watts et al. (2018) challenge this pattern. They argue that the original longitudinal associations found by Mischel and his team were based on small and highly selective samples of children whose parents were highly qualified academics. Their conceptual replication instead uses a larger and more diverse sample of children, i.e. a sample that also includes children from less thriving backgrounds. As noted before, their results suggest for smaller effects of children's waiting times on behavioral outcomes at age 15 and that these effects vanish as soon as additional background factors are accounted for.

Yet, other replications and adaptations of the initial study reconfirm the relevance of self-regulation. For example, individuals who were more patient as child, commit less crimes until and in adulthood (Akerlund et al., 2016; Moffitt et al., 2011). They also have a lower BMI (Sutter et al., 2015; Bub et al., 2016; Golsteyn et al., 2014), perform better financially (Moffitt et al., 2011;

[4] This experimental literature is yet in its infancy. It for example adresses whether changes in the default choice setting can moderate self-regulation behavior (Carroll et al., 2009; Sutter et al., 2015).

[5] Sethi et al. (2000) show that the onset of differentials seems to be already observable in children as young as 18 months: children who are better at coping with a brief absence of their mother also perform better on the Marshmallow test at the age of 5 years.

Golsteyn et al., 2014), depend on substances less frequently (Moffitt et al., 2011), are somewhat less likely to smoke (Fuchs, 1982; Bickel et al., 1999), and are healthier in general (Bub et al., 2016; Moffitt et al., 2011).

A potential pathway of the relationship between time preferences and lifetime outcomes may be through educational attainment. Studies from the original Marshmallow tests found patient children to be rated more favorable by their parents in terms of competence, attentiveness, and their ability do deal with frustration and stress (Mischel et al., 1988), and to perform better in school (Mischel et al., 1989).

More recent studies indicate that impatience relates to more disruptive behavior in school (Castillo et al., 2011), decreases the probability of graduating from high school (Castillo et al., forthcoming), or increases drop-out from college (Cadena and Keys, 2015). Benjamin et al. (2013) further report that patient children achieve higher Scholastic Aptitude Test (SAT) scores. Bettinger and Slonim (2007), on the other hand, do not find a correlation of time preferences with school performance.[6] Complementing the link between (im)patience and education, Golsteyn et al. (2014) find that the effect of time preferences on lifetime outcomes falls substantially if they account for educational attainment. Controlling for ability reduces the estimates as well, though to a lesser extent.

Further research demonstrates a clear, positive relationship between patience and cognitive skills. There is, however, only little research yet that examines whether this relationship is causal indeed, and in which direction causality works. For Chilean high-schools students, Benjamin et al. (2013) not only report on a link between cognitive skills and time preferences, but they suggest a possible causal impact of cognitive resources on expressed preferences. Correlations between time preferences and cognitive abilities are also found in adult populations (Shamosh and Gray, 2008; Dohmen et al., 2010), but it is unclear for these studies, which trait begets which.

Finally, and as mentioned before, we enhance the study of Lorenz et al. (2016), who conduct a cross-sectional analysis using the kindergarten cohort of the NEPS data. Their results imply a positive relation between children's patience and mathematical, language, and cognitive skills as well as working memory, even when controlling for social background. We extend their approach by exploring the longitudinal dimension of the data in order to examine whether early self-regulation differentials add to mathematical competence development.

2.3 Data and Empirical Strategy

2.3.1 Data

For our analyses we use data from the kindergarten children cohort of the German National Educational Panel Study (NEPS, Starting Cohort 2) (Blossfeld et al., 2011). A sample of four year old preschoolers attending kindergarten in Germany was first surveyed in 2011 and has since been followed into primary school and beyond. In the first wave, roughly 3,000 children took part in the study, but only 576 children could be followed into school which leads to a substantial decline of suitable observations for our analyses. Because we restrict our data to a balanced panel, and because of missing values in key variables our analysis sample further decreases to 370 observations.[7]

---

6 For college students, evidence is mixed as well inasmuch as impatient students do not exert significantly less effort, but perform less well in exams (Non and Tempelaar, 2016). On top of that, further research suggests that time preferences may still be malleable in early adulthood and that education itself has an impact on it (Perez-Arce, 2017).

7 Table A.1 in the appendix shows that the majority of variables does not differ significantly between the analysis sample and the full cohort sample in terms of normalized differences (Imbens and Wooldridge, 2009). Initial math

In each wave every child was tested in various competence domains over two consecutive days. The assessments were conducted individually in kindergarten and in groups in primary school. In 2012, i.e. the second wave, when the preschoolers were around their sixth birthday, their self-regulatory abilities were measured with the following test: Each child was shown a small bag with unknown content at the end of the first day of testing. The child was told that there were presents inside and it was then offered the choice to either draw one present from the bag immediately or two presents on the next day. After making sure that the child understood the implications of the decision, it was asked to choose between the two options.

Although Mischel's Marshmallow test inspired the NEPS-test of self-regulatory abilities, the two procedures differ: Most importantly, the children in Mischel's experiments knew what kind of gratification they would get and were in most cases exposed to it while waiting.[8] This is an implementation of what Neubauer et al. (2011) call the waiting paradigm (e. g. Mischel and Metzner, 1962; Langenfeld et al., 1997; Mischel et al., 1988; Shoda et al., 1990). Mischel and Ebbesen (1970) and Mischel et al. (1972) show that waiting times for the preferred but delayed reward reduce dramatically if children direct their attention towards the rewards (e. g. if the reward is placed directly in front of them instead of being out of their sight). In the NEPS-test, the children did not know what kind of present they could expect and were not exposed to it during the waiting period. Therefore, the NEPS-test is an implementation of what Neubauer et al. (2011) call the choice paradigm (e. g. Mischel and Gilligan, 1964; Bochner and David, 1968). According to Lemmon and Moore (2007), such tests are valid measures for children's delay of gratification from the age of four years on.

The NEPS also provides a set of competence measures to assess children's mathematical, language, and cognitive skills as well as their working memory. Mathematical competence, however, is the only competence measure that was assessed in 2012 and in the following years. We therefore focus on the mathematical competence domain as dependent variable because we are especially interested in competence development and because it is a major predictor of educational attainment (Watts et al., 2014, 2017).

The mathematical competence test procedure requires that children at the initial age of our target population (5-6 years) have already developed an understanding of the concept of numbers and are able to answer simple questions about comparisons of sets, counting tasks or ordinal aspects with the aid of illustrative materials[9] (Neumann et al., 2013). To ensure that mathematical competence is measured independently from reading competence, the items were read to the children and the children answered using pictures or arabic numbers smaller than 20 (Leibniz Institute for Educational Trajectories, 2015, p. 5). Based on such tests, the scientific use file of the NEPS provides weighted maximum likelihood estimates (WLE) of the observed responses as measure of children's mathematical competence. In order to enable comparisons over time, the competence scores were linked in a scaling study between kindergarten and grade 1 (Schnittjer, 2018) and, using anchor items, between grade 1 and 2 (Schnittjer and Gerken, 2018).

Cognitive basic skills were also tested in the second wave by assessing perceptual speed and reasoning abilities. These skills do not depend on domain-specific cognitive processes, such as language skills, but are general abilities, and core elements of the so-called fluid intelligence, which represents an important determinant of learning processes (Primi et al., 2010). To measure perceptual speed the participants have to match figures with graphical symbols as quickly as

competence, however, is higher for children who participated in the survey in all three waves which are of interest to us. We are therefore quite reserved about generalizing our findings.

8 Mischel and his co-authors experimented with a variety of different experimental settings. The one sketched here is probably the most widely known implementation.

9 For example: "In this bowl are four stones. Now I add three stones. [The bowl is covered, so the child cannot see what is inside.] Can you tell me, how many stones are in the bowl now?" (Schnittjer and Duchhardt, 2015, p. 3 ; our translation).

possible. For reasoning, a geometrical element has to be selected which fits the logical rules of a shown pattern of such elements (Haberkorn and Pohl, 2013). In the scientific use file of NEPS, the results of these tests are available as sum scores of correct answers.

In addition to competence measures and child characteristics, the NEPS provides information on family background. In our case, we can use context data for a rich set of cross-sectional as well as longitudinal information on background characteristics of the children and their families.[10]

2.3.2 Empirical strategy

To assess the impact of the decision to wait on the child's math competence levels, we estimate the following model:

$$M_{i,t} = \alpha \cdot delay_{i,t=k} + \mathbf{X_i}\beta + \epsilon_i, \tag{2.1}$$

where $M_{i,t}$ is the mathematical competence of child i in time period $t \in k, 1, 2$ (kindergarten, grade 1 or grade 2), $delay$ is a dummy variable indicating whether the child decided to wait in the delay of gratification task, so that α is the coefficient of interest. $\mathbf{X_i}$ is a set of individual background characteristics as outlined in detail later on, which first excludes and later includes measures for basic cognitive skills in $t = k$; ϵ_i is the individual error term clustered at kindergarten group level.

We next examine whether the child's decision to wait also relates to the gains in skills over time in a second set of estimations:

$$M_{i,t} - M_{i,t-s} = \alpha \cdot delay_{i,t=k} + \mathbf{X_i}\beta + \epsilon_i, \tag{2.2}$$

i. e. we measure the effect of being able to wait on the development of mathematical competence. $M_{i,t}$ displays math competence in $t = 1$ or $t = 2$ and $M_{i,t-s}$ is math competence one or two periods earlier ($t = k$ or $t = 1$).

In a final step, to assess potential effect heterogeneity within the initial mathematical competence distribution, we add a dummy for whether the child's mathematical competence was below average in kindergarten ($D_{i,M_{i,k}<\bar{M_k}}$) and interact it with delay of gratification:

$$M_{i,t} - M_{i,t-s} = \alpha \cdot delay_{i,t=k} + \delta \cdot (D_{i,M_{i,k}<\bar{M_k}} \cdot delay_{i,t=k}) + \gamma \cdot D_{i,M_{i,k}<\bar{M_k}} + \mathbf{X_i}\beta + \epsilon_i. \tag{2.3}$$

Based on prior research, the vector $\mathbf{X_i}$ contains a range of covariates to account for likely influences on both the child's competence development and his or her self-regulation. In particular, we control for the following child's characteristics: age, gender, and whether it lives in East or West Germany.

To account for a potential confounding impact of the child's personality on self-regulation, we include parental ratings of the child's Big Five personality traits, i. e. openness, conscientiousness, extraversion, agreeableness, and neuroticism (McCrae and John, 1992).[11]

Parental background is controlled for by including covariates on migration background, whether the interviewed parent lives with a partner, parental education, and household income. We further account for the learning environment at home by controlling for the number of books at home, as well as the number of siblings.[12]

---

[10] Table A.2 in the appendix provides information on all the variables we use in our analyses.

[11] Parental ratings of children's personality were measured in a specifically designed questionnaire by Müller et al. (2016).

[12] In additional specifications, we included further context information on kindergarten characteristics. The additional estimations included children-to-kindergarten-staff ratio as a rough global indicator for childcare quality,

Table 2.1: Descriptive Statistics

	Pooled		Patient		Impatient		Difference	
	Mean	(S.D.)	Mean	(S.D.)	Mean	(S.D.)	Diff	(p-value)
Delayed gratification (DG)	0.39	(0.49)						
Competence measures								
Math competence: kindergarten	0.49	(0.99)	0.74	(0.95)	0.33	(0.99)	0.41***	(0.00)
Math competence: grade 1	1.80	(1.13)	2.04	(1.02)	1.64	(1.17)	0.40***	(0.00)
Math competence: grade 2	2.51	(1.14)	2.77	(1.02)	2.35	(1.18)	0.42***	(0.00)
Perceptual speed	18.88	(5.56)	19.57	(6.11)	18.44	(5.14)	1.13*	(0.07)
Reasoning	5.82	(2.46)	6.19	(2.41)	5.58	(2.47)	0.60**	(0.02)
Child characteristics								
East German	0.21	(0.41)	0.15	(0.36)	0.24	(0.43)	−0.09**	(0.03)
Male child	0.49	(0.50)	0.47	(0.50)	0.50	(0.50)	−0.03	(0.60)
Age in months	71.15	(3.76)	71.35	(3.77)	71.03	(3.75)	0.32	(0.42)
Child's personality								
Big Five: Extraversion	8.07	(1.67)	7.95	(1.70)	8.15	(1.65)	−0.19	(0.28)
Big Five: Conscientiousness	6.31	(1.58)	6.39	(1.64)	6.25	(1.54)	0.14	(0.42)
Big Five: Agreeableness	5.90	(1.66)	6.13	(1.57)	5.75	(1.70)	0.39**	(0.03)
Big Five: Openness/Intellect	8.26	(1.24)	8.43	(1.30)	8.15	(1.20)	0.28**	(0.04)
Big Five: Neuroticism	3.60	(1.86)	3.53	(1.95)	3.64	(1.80)	−0.12	(0.57)
Parental background								
Migration background	0.07	(0.26)	0.10	(0.31)	0.05	(0.22)	0.05*	(0.09)
Highest CASMIN:								
Basic sec. educ. or less	0.04	(0.20)	0.03	(0.16)	0.05	(0.22)	−0.02	(0.29)
Intermediate sec. educ.	0.34	(0.47)	0.29	(0.46)	0.36	(0.48)	−0.07	(0.15)
Univ. entrance qualif. or more	0.62	(0.48)	0.68	(0.47)	0.59	(0.49)	0.09*	(0.07)
Household income	0.09	(0.66)	0.09	(0.63)	0.10	(0.68)	−0.01	(0.86)
Living together with a partner	0.93	(0.25)	0.94	(0.23)	0.92	(0.26)	0.02	(0.45)
Home environment								
Number of siblings	1.04	(0.88)	1.06	(0.78)	1.02	(0.93)	0.03	(0.71)
More than 100 books at home	0.62	(0.49)	0.65	(0.48)	0.60	(0.49)	0.04	(0.39)
Observations	370		144		226		370	

Notes: Data: NEPS SC2 5.1.0, own calculations. Difference displays the difference between patient and impatient individuals. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

2.4 Results

We start by presenting descriptive differences between children who decide to wait and who do not. In a next step we discuss our baseline specification OLS models with mathematical competence levels as dependent variable. Note that we always cluster our standard errors at the kindergarten group level to control for within-group error correlation (Angrist and Pischke, 2009).[13] We first estimate level differences in math competence by self-regulation in kindergarten, grade 1 and grade 2 separately. We then analyze how delayed gratification is related to the competence development of children by using gains in mathematical competence as dependent variable. Finally, we interact initial mathematical competence with the decision to delay gratification to detect whether children, who were initially weaker in the math test, show different competence gains over time.

2.4.1 Descriptive differences between patient and impatient children

Differences in test scores and characteristics between patient and impatient children are reported in table 2.1. In our sample, 39% of the children decided to wait for the next day in order to receive two presents instead of one present they could have got immediately. We further see a strong assocation between children's ability to wait and their mathematical competence: Patient children outperform impatient children in all domains.

There, however, are not many statistical differences in child characteristics. On average, patient children score higher on the Big Five measure of Openness to Experience and Agreeableness, come from a household where parents have higher educational attainment, live less often in Eastern Germany, and more often have a migration background than impatient children. They however do not statistically differ in terms of age, gender, the Big Five traits other than Openness and Agreeableness, and the learning environment at home.

2.4.2 Self-regulation and competence levels

The purely descriptive patterns suggest a strong positive association between patience and mathematical competence. To net out a confounding impact of the child's characteristics, we next analyze the association between the ability to wait in kindergarten and mathematical competence in all observed years in a regression framework. That is, we condition on the covariates as described before and run multiple regressions for the kindergarten wave, where both mathematical competences and self-regulation were measured, as well as for grade 1 and grade 2 for which we examine levels of and gains in mathematical competences. For each wave, we regress two specifications, one accounting for self-regulation only and another that additionally includes general cognitive abilities in order to capture potentially confounding effects. The main results of these regressions are summarized in table 2.2, full results are given in the appendix, table A.3.

The results in columns 1 and 2 of table 2 show the cross-sectional relationship between patience in kindergarten and mathematical competence. Both constructs were measured on the same day, so that the results cannot be interpreted as causal. The coefficients show a strong positive relationship between the decision to wait and mathematical competence in kindergarten. With a competence score differential of 0.31 points (roughly 31% of a standard deviation), the size of the level difference is substantial (column 1). When additionally controlling for general cognitive ability in kindergarten (column 2), the differential decreases only slightly, implying that

group size, or gender composition. Because of large unit non-response at the kindergarten management level, sample size is substantially lower. This yields trivial results which are not reported here.

[13] Our results are, however, not sensitive to this as we show in section 2.5.

Table 2.2: Effects on Math competence level

	Kindergarten		Grade 1		Grade 2	
	(1)	(2)	(3)	(4)	(5)	(6)
Delayed gratification (DG)	0.317***	0.233***	0.299***	0.217**	0.313***	0.222**
	(0.089)	(0.085)	(0.104)	(0.103)	(0.103)	(0.096)
Perceptual speed std		0.243***		0.293***		0.218***
		(0.048)		(0.056)		(0.063)
Reasoning std		0.254***		0.205***		0.310***
		(0.044)		(0.051)		(0.060)
N	370	370	370	370	370	370
adj. R^2	0.199	0.331	0.131	0.229	0.153	0.271

Notes: Data: NEPS SC2 5.1.0, own calculations. All estimations contain a constant and all other explanatory variables named in table 2.1. For results on controls see appendix, table A.3. Standard errors, reported in parentheses, are clustered at kindergarten group level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

general cognitive skills are a confounding factor, yet that self-regulation is not fully determined by or simply representing these skills.[14]

In columns 3 to 6, we present the results of the decision to delay gratification in kindergarten on mathematical competence in grade 1 and 2, again based on specifications ex- or including general cognitive skills. The magnitude of the level differences as given in the kindergarten wave remains almost unchanged.[15]

Our results indicate a substantial positive relationship between children's ability to delay gratification and current as well as future mathematical competence. In terms of inequalities we see a competence gap between more and less patient children already in our first wave, i.e. when patience is measured, and that it persists over the following two years, even if initial general cognitive ability is controlled for.

2.4.3 Effects on competence development

We next examine whether the competence gap between patient and impatient children persists and estimate a value-added-type specification. The dependent variable in this setting is the difference in mathematical competence between two waves. We consider three different time frames and examine changes in mathematical competences: from kindergarten to grade 1 (table 2.3, columns 1 and 2), from kindergarten to grade 2 (columns 3 and 4) and, finally, changes between grade 1 and grade 2 (columns 5 and 6).[16] We again estimate two sets of specifications, first accounting only for socio-demographic characteristics and adding general cognitive ability in the second set of models.

The results imply that self-regulation has no impact on the change in mathematical competence in the first two years of primary school. Adding general cognitive ability, the coefficients are again slightly attenuated and in general do not suggest that general cognitive abilities impact

[14] Note however that explained variation increases substantially if general cognitive skills are accounted for. For the other covariates, we observe that being male, age, being open for experiences, and higher parental education are positively related to math competence, while being neurotic, extraverted and having a migration background are negatively related to kindergarten math competence (cf. A.3, column 2).

[15] Similar to the estimates in column 2, the explanatory power of the model increases substantially if general cognitive abilities are accounted for.

[16] Full results are displayed in table A.4 in the appendix.

Table 2.3: Math competence development

	Kindergarten - Grade 1		Kindergarten - Grade 2		Grade 1 - Grade 2	
	(1)	(2)	(3)	(4)	(5)	(6)
Delayed gratification (DG)	−0.003	−0.003	0.039	0.028	0.037	0.025
	(0.100)	(0.100)	(0.095)	(0.094)	(0.099)	(0.098)
Perceptual speed std		0.053		0.013		−0.037
		(0.065)		(0.069)		(0.058)
Reasoning std		−0.043		0.052		0.089*
		(0.047)		(0.058)		(0.048)
N	370	370	370	370	370	370
adj. R^2	0.002	0.000	0.037	0.035	0.056	0.060

Notes: Data: NEPS SC2 5.1.0, own calculations. All estimations contain a constant, control for months between tests and all other explanatory variables named in table 2.1. For results on controls see appendix, table A.4. Standard errors, reported in parentheses, are clustered at kindergarten group level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

competence gains, except for the development between grade 1 and 2, where the coefficient for reasoning is statistically different from zero.

The findings shown in table 2.2 and 2.3 together suggest that despite the relation between self-regulation measured in kindergarten and children's math competence level, self-regulation does on average not affect competence development in the first two years of primary school. Put differently, we observe a gap in math competence between patient and impatient children which already exists in kindergarten and persists until grade 2, but it neither widens nor narrows because of children's ability to wait.[17]

2.4.4 Heterogeneity analyses

We have seen for the full sample, that children's patience does not affect their mathematical competence development. The ability to delay gratification might however—via dynamic cross-complementarity—be differently useful for children of different initial competence endowment. Children with lower initial mathematical competence may particularly benefit from higher self-regulation.

To examine potential effect heterogeneities with respect to the initial level of math competence, we run estimations according to the model outlined in equation 3, i.e. we add a dummy indicating whether the child's math competence was below average in kindergarten and, in a second step, by interacting this dummy with the decision to wait. The results in table 4 show that, compared to the findings in table 3, the coefficients for delayed gratification slightly increase in the first step (columns 1, 3, and 5), but remain statistically insignificant.

We do, however, see that children with low initial skills exhibit larger competence gains: The coefficients for the dummy variables on low initial math competence are rather large and statistically different from zero.

The results of the interaction term further show that for the development between kindergarten and grade 1 patient children with low initial math competence gain more than patient children with high initial math competence. They also gain more math competence than impa-

[17] This conflicts with expectations from the model of skill formation. As outlined before, the NEPS data do, however, not provide details on children's investments, e.g. in terms of time spent on homework or learning, so that more in-depth analyses are not feasible.

Table 2.4: Effect heterogeneity by initial math competence

	Kindergarten - Grade 1		Kindergarten - Grade 2		Grade 1 - Grade 2	
	(1)	(2)	(3)	(4)	(5)	(6)
Delayed gratification (DG)	0.071	−0.017	0.081	0.030	0.083	0.077
	(0.097)	(0.110)	(0.093)	(0.105)	(0.094)	(0.144)
Math comp. below avg.	0.521***	0.412***	0.383***	0.319**	0.606***	0.601***
	(0.111)	(0.119)	(0.122)	(0.143)	(0.095)	(0.116)
DG × Math comp. below avg.		0.416*		0.244		0.013
		(0.217)		(0.211)		(0.208)
N	370	370	370	370	370	370
adj. R^2	0.049	0.054	0.059	0.059	0.147	0.144
Wald test: p-value[a]		0.036		0.143		0.501

Notes: Data: NEPS SC2 5.1.0, own calculations. All estimations contain a constant, control for months between tests and all other explanatory variables named in table 2.1. For results on controls see appendix, table A.5. Standard errors, reported in parentheses, are clustered at kindergarten group level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.
[a] Test of the hypothesis H_0: Delayed gratification (DG) + DG × Math comp. below avg. = 0.

tient children, as the p-values from the Wald-test implies. There is, however, no advantage for patient children between kindergarten and grade 2 or between grades 1 and 2.

2.5 Robustness checks

This section details the robustness checks we conducted. We first examined whether clustering at different levels than at the kindergarten level, i.e. not at all, or at grade 1 level, plays a role. We then run additional models using grade 1 or grade 2 covariates, and finally checked whether not accounting for children's personality traits matters.

Different standard error calculations: We estimated our models in section 2.4 with standard errors clustered at the kindergarten group level in all of our specifications to account for unobservable group composition or environment. As this most likely changes when the child leaves kindergarten and becomes a student in primary school, we run additional estimates with clustered standard errors at the classroom level. The standard errors, however, change only marginally in these analyses or, as additional exercise, if we do not cluster at all (see table A.6 in the appendix).

Using grade 1 controls: When examining the effects of delayed gratification on math competence development from grade 1 to grade 2, we used the information on the child's sociodemographics from the kindergarten wave as control variables. It is however possible that children's circumstances changed between kindergarten and grade 1. We therefore also estimated the math competence development from grade 1 to grade 2, accounting for all covariates as outlined before, from the grade 1 wave. Because of missing data in some of the background information the sample decreases to 221 observations.[18] The results, however, change only marginally (see table A.7 in the appendix).

Personality measures as controls: As noted before, we control for the child's Big Five personality traits to account for potential confounding relationships to self-regulation. The measures in our main models are based on parental ratings, surveyed in the first wave, i.e. when the child

[18] We also considered running random effects models, but the feasible longitudinal sample was way too small to turn this into a meaningful endeavor.

was attending kindergarten. Personality, however, still evolves in this age group (Herzhoff et al., 2017) and may as well be related to changes in self-regulation. The latter is not available in the NEPS and the child's Big Five personality traits are re-measured only in grade 2.[19]

Becker et al. (2012) furthermore suggest that facets of individuals' personality are related to patience or time-discounting[20] but that they are complements in explaining lifetime outcomes. The self-regulation test in the NEPS data, however, differs from the typical time preference measurements in economics. Therefore, as an additional robustness check, we estimated our models without controlling for the Big Five personality traits. The exclusion of the Big Five personality traits only marginally changes the coefficients for competence levels, competence development, and effect heterogeneity (cf. table A.8, table A.9, and table A.10, respectively in the appendix).

2.6 Conclusion

We contribute to the literature on early life skills differentials and the role of self-regulation in this. To do so, we examined how children's ability to delay gratification relates to their mathematical competence and its development. We use NEPS data and find that, even when controlling for general cognitive skills, there is a positive relationship between the ability to wait in kindergarten and mathematical competence from kindergarten through grade two of primary school. The relationship is quite strong with patience explaining 20 to 30% of a standard deviation in mathematical competence. Furthermore, the estimates for the level differences do not decrease substantially over the first years of primary school.

In a second step we examined the effect of kindergarten patience on math competence gains. Complementing the level differentials, we do not find that patience affects the speed of competence gains, but heterogeneity analyses suggest that being patient in kindergarten positively affects mathematical competence gains at the transition from kindergarten to the first year of primary school for children with lower initial mathematical competence. Self-regulation, however, seems to play no further role for competence development between grade 1 and grade 2.

In the NEPS data, information on children's self-regulation is given by observable behavior, and their mathematical competences are derived from specifically developed assessments tests. Both sets of indicators are therefore more reliable than e. g. self-reported data or grades, which strengthens our results. We, however, are not able to draw straightforward causal claims from our analyses, because we have no exogenous variation in the NEPS data. Future research could therefore attempt to establish more evidence on causality by, for example, using interventions designed to foster children's self-regulation skills. This would help to derive policy implications on how to decrease the competence gap that relates to differences in children's patience or self-regulation.

[19] Although pairwise correlations between a child's self-regulation and the Big Five personality traits do not indicate substantial changes in the relation between kindergarten and grade 2, using grade 2 data induces yet another decrease in sample size, so that interpretation gets problematic.

[20] Interestingly, there is yet only some small corpus of research addressing the relation between the Big Five personality traits and measures of self-control (Hoyle and Davisson, 2016; Becker et al., 2012).

Appendix

Table A.1: Comparison of full and analysis sample

	Full sample		Dropout sample		Analysis sample		Difference		
	Mean	(S.D.)	Mean	(S.D.)	Mean	(S.D.)	Diff	(p-value)	Norm Diff
Delayed gratification (DG)	0.35	(0.48)	0.35	(0.48)	0.39	(0.49)	−0.04	(0.11)	0.06
Competence measures									
Math competence: kindergarten	0.01	(1.17)	−0.07	(1.18)	0.49	(0.99)	−0.56***	(0.00)	0.36
Math competence: grade 1	1.75	(1.17)	1.61	(1.28)	1.80	(1.13)	−0.18	(0.13)	0.11
Math competence: grade 2	2.44	(1.18)	2.24	(1.27)	2.51	(1.14)	−0.27**	(0.03)	0.16
Perceptual speed	17.84	(6.09)	17.67	(6.16)	18.88	(5.56)	−1.21***	(0.00)	0.15
Reasoning	5.32	(2.38)	5.24	(2.36)	5.82	(2.46)	−0.58***	(0.00)	0.17
Child characteristics									
East German	0.22	(0.41)	0.22	(0.41)	0.21	(0.41)	0.01	(0.58)	−0.02
Male child	0.50	(0.50)	0.50	(0.50)	0.49	(0.50)	0.01	(0.71)	−0.02
Age in months	70.69	(3.94)	70.62	(3.96)	71.15	(3.76)	−0.53**	(0.01)	0.10
Child's personality									
Big Five: Extraversion	8.09	(1.70)	8.09	(1.71)	8.07	(1.67)	0.02	(0.83)	−0.01
Big Five: Conscientiousness	6.20	(1.70)	6.17	(1.73)	6.31	(1.58)	−0.14	(0.14)	0.06
Big Five: Agreeableness	5.80	(1.71)	5.78	(1.72)	5.90	(1.66)	−0.12	(0.24)	0.05
Big Five: Openness/Intellect	8.18	(1.38)	8.16	(1.42)	8.26	(1.24)	−0.10	(0.18)	0.05
Big Five: Neuroticism	3.60	(1.82)	3.59	(1.82)	3.60	(1.86)	−0.00	(0.97)	0.00
Parental background									
Migration background	0.15	(0.36)	0.17	(0.37)	0.07	(0.26)	0.10***	(0.00)	−0.21
Highest CASMIN:									
Basic sec. educ. or less	0.09	(0.29)	0.11	(0.31)	0.04	(0.20)	0.07***	(0.00)	−0.18
Intermediate sec. educ.	0.33	(0.47)	0.33	(0.47)	0.34	(0.47)	−0.00	(0.88)	0.01
Univ. entrance qualif. or more	0.58	(0.49)	0.56	(0.50)	0.62	(0.48)	−0.06**	(0.03)	0.09
Household income	0.02	(1.02)	−0.01	(1.10)	0.09	(0.66)	−0.10**	(0.03)	0.08
Living together with a partner	0.90	(0.30)	0.89	(0.31)	0.93	(0.25)	−0.04**	(0.01)	0.10
Home environment									
Number of siblings	1.08	(0.92)	1.09	(0.93)	1.04	(0.88)	0.05	(0.31)	−0.04
More than 100 books at home	0.55	(0.50)	0.53	(0.50)	0.62	(0.49)	−0.09***	(0.00)	0.12
Observations	2644		2274		370		2644		2644

Notes: Data: NEPS SUF, SC2 5.1.0, own calculations. The full sample contains all individuals for whom we observe data on delayed gratification and kindergarten math competence; all other variables have fewer observations than stated in the full and dropout sample. Difference displays the difference between analysis and dropout sample. Norm Diff displays normalized differences as suggested by Imbens and Wooldridge (2009) where the critical value typically is 0.25 or -0.25. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A.2: Variable definitions

Variable	Definition
Delayed gratification	Dummy equal to one if the child decided to wait for the second gift
Competence measures	
Math competence	WLE score of child's math competence
Perceptual speed	Sum score of child's perceptual speed
Reasoning	Sum score of child's reasoning abilities
Months between tests	Number of months between the two survey dates
Child demographics	
Male child	Dummy equal to one if the child is male
Age (months)	Child's age in months
East German	Dummy equal to one if interviewed parent lives in East Germany
Child personality	
Big Five: Extraversion std	Parental report z-standardized over full NEPS kindergarten sample
Big Five: Conscientiousness std	Parental report z-standardized over full NEPS kindergarten sample
Big Five: Agreeableness std	Parental report z-standardized over full NEPS kindergarten sample
Big Five: Openness/Intellect std	Parental report z-standardized over full NEPS kindergarten sample
Big Five: Neuroticism std	Parental report z-standardized over full NEPS kindergarten sample
Parental background	
Migration background	Dummy equal to one if at least one parent and both parents of the other parent are born abroad
Living together with a partner	Dummy equal to one if the interviewed parent lives with a partner
CASMIN	Highest educational level of the parents living in the same household with the child coded using the CASMIN classification
Home environment	
Books at home: more than 100	Dummy equal to one if more than 100 books are available in parental home
Number of siblings	Number of siblings living in the same household with the child

Table A.3: Effects on Math competence level

	Kindergarten		Grade 1		Grade 2	
	(1)	(2)	(3)	(4)	(5)	(6)
Delayed gratification (DG)	0.317***	0.233***	0.299***	0.217**	0.313***	0.222**
	(0.089)	(0.085)	(0.104)	(0.103)	(0.103)	(0.096)
Perceptual speed std		0.243***		0.293***		0.218***
		(0.048)		(0.056)		(0.063)
Reasoning std		0.254***		0.205***		0.310***
		(0.044)		(0.051)		(0.060)
East German	−0.088	−0.014	−0.190	−0.100	−0.269*	−0.202
	(0.154)	(0.141)	(0.140)	(0.132)	(0.153)	(0.150)
Male child	0.143	0.222**	0.203*	0.276**	0.183	0.270**
	(0.096)	(0.086)	(0.109)	(0.107)	(0.115)	(0.119)
Age in months std	0.777***	0.527***	0.499**	0.236	0.324	0.070
	(0.206)	(0.197)	(0.228)	(0.214)	(0.230)	(0.221)
Extraversion std	−0.139***	−0.128**	−0.120	−0.114	−0.068	−0.051
	(0.053)	(0.052)	(0.075)	(0.072)	(0.069)	(0.070)
Conscientiousness std	−0.001	−0.070	−0.030	−0.100	0.054	−0.019
	(0.055)	(0.053)	(0.064)	(0.063)	(0.061)	(0.060)
Agreeableness std	0.010	0.003	0.034	0.034	0.001	−0.011
	(0.045)	(0.045)	(0.062)	(0.062)	(0.056)	(0.051)
Openness/Intellect std	0.207***	0.192***	0.093	0.082	0.215***	0.196***
	(0.061)	(0.056)	(0.088)	(0.080)	(0.071)	(0.066)
Neuroticism std	−0.093*	−0.086*	−0.119*	−0.111*	0.036	0.044
	(0.055)	(0.048)	(0.070)	(0.067)	(0.064)	(0.060)
Migration background	−0.569***	−0.554***	−0.388**	−0.384**	−0.480**	−0.455***
	(0.200)	(0.191)	(0.191)	(0.181)	(0.188)	(0.149)
CASMIN (ref. *Basic sec. educ. or less*)						
– Intermediate sec. educ.	0.444	0.535**	0.875***	0.981***	0.445	0.531*
	(0.290)	(0.258)	(0.327)	(0.284)	(0.334)	(0.270)
– Univ. entrance qualif. or more	0.711**	0.780***	1.191***	1.286***	0.794**	0.848***
	(0.299)	(0.272)	(0.341)	(0.298)	(0.342)	(0.282)
More than 100 books at home	0.157	0.098	0.208*	0.140	0.261**	0.208*
	(0.110)	(0.103)	(0.121)	(0.120)	(0.119)	(0.121)
Household income	0.047	0.014	0.021	−0.014	0.059	0.025
	(0.082)	(0.067)	(0.109)	(0.095)	(0.104)	(0.090)
Living together with a partner	0.086	0.081	−0.037	−0.041	0.067	0.060
	(0.228)	(0.204)	(0.261)	(0.234)	(0.252)	(0.244)
Number of siblings	−0.097*	−0.035	−0.098	−0.031	−0.063	−0.002
	(0.053)	(0.045)	(0.062)	(0.059)	(0.069)	(0.068)
Constant	0.478	0.011	1.134**	0.633	1.903***	1.434***
	(0.426)	(0.353)	(0.465)	(0.398)	(0.456)	(0.395)
N	370	370	370	370	370	370
adj. R^2	0.199	0.331	0.131	0.229	0.153	0.271

Notes: Data: NEPS SC2 5.1.0, own calculations. Standard errors, reported in parentheses, are clustered at kindergarten group level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A.4: Effects on Math competence development (full table)

	Kindergarten - Grade 1		Kindergarten - Grade 2		Grade 1 - Grade 2	
	(1)	(2)	(3)	(4)	(5)	(6)
Delayed gratification (DG)	−0.003	−0.003	0.039	0.028	0.037	0.025
	(0.100)	(0.100)	(0.095)	(0.094)	(0.099)	(0.098)
Perceptual speed std		0.053		0.013		−0.037
		(0.065)		(0.069)		(0.058)
Reasoning std		−0.043		0.052		0.089*
		(0.047)		(0.058)		(0.048)
Months between tests	0.118*	0.116*	0.191***	0.194***	0.234***	0.221***
	(0.063)	(0.063)	(0.055)	(0.054)	(0.061)	(0.060)
East German	−0.114	−0.098	−0.210	−0.206	−0.089	−0.100
	(0.151)	(0.150)	(0.133)	(0.135)	(0.121)	(0.122)
Male child	0.045	0.041	0.044	0.056	0.014	0.028
	(0.093)	(0.095)	(0.104)	(0.108)	(0.104)	(0.108)
Age in months std	−0.186	−0.204	−0.344*	−0.370*	−0.225	−0.232
	(0.195)	(0.193)	(0.199)	(0.204)	(0.186)	(0.188)
Extraversion std	0.018	0.012	0.062	0.066	0.044	0.053
	(0.065)	(0.064)	(0.062)	(0.063)	(0.060)	(0.062)
Conscientiousness std	−0.034	−0.036	0.043	0.034	0.080	0.074
	(0.060)	(0.061)	(0.062)	(0.061)	(0.055)	(0.055)
Agreeableness std	0.019	0.025	−0.029	−0.032	−0.048	−0.055
	(0.058)	(0.059)	(0.052)	(0.050)	(0.054)	(0.054)
Openness/Intellect std	−0.117	−0.113	0.007	0.003	0.126*	0.120
	(0.087)	(0.086)	(0.062)	(0.062)	(0.075)	(0.075)
Neuroticism std	−0.025	−0.024	0.124**	0.124**	0.147***	0.147***
	(0.052)	(0.051)	(0.051)	(0.052)	(0.053)	(0.053)
Migration background	0.168	0.157	0.094	0.101	−0.059	−0.045
	(0.235)	(0.238)	(0.222)	(0.220)	(0.178)	(0.174)
CASMIN (ref. *Basic sec. educ. or less*)						
Intermediate sec. educ.	0.376	0.392	−0.086	−0.080	−0.426	−0.435
	(0.286)	(0.287)	(0.303)	(0.301)	(0.266)	(0.265)
Univ. entrance qualif. or more	0.426	0.451	0.028	0.027	−0.355	−0.381
	(0.300)	(0.302)	(0.295)	(0.294)	(0.266)	(0.267)
More than 100 books at home	0.055	0.044	0.075	0.071	0.009	0.018
	(0.113)	(0.116)	(0.113)	(0.119)	(0.107)	(0.105)
Household income	−0.029	−0.031	0.015	0.011	0.047	0.046
	(0.082)	(0.082)	(0.078)	(0.078)	(0.058)	(0.060)
Living together with a partner	−0.106	−0.104	−0.009	−0.010	0.081	0.080
	(0.230)	(0.228)	(0.178)	(0.181)	(0.204)	(0.205)
Number of siblings	−0.009	−0.003	0.017	0.023	0.031	0.031
	(0.056)	(0.054)	(0.058)	(0.057)	(0.063)	(0.063)
Constant	−0.662	−0.689	−2.450**	−2.559**	−1.343**	−1.238*
	(0.895)	(0.885)	(1.163)	(1.155)	(0.632)	(0.635)
N	370	370	370	370	370	370
adj. R^2	0.002	0.000	0.037	0.035	0.056	0.060

Notes: Data: NEPS SC2 5.1.0, own calculations. Standard errors, reported in parentheses, are clustered at kindergarten group level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A.5: Effect heterogeneity by initial math competence

	Kindergarten - Grade 1		Kindergarten - Grade 2		Grade 1 - Grade 2	
	(1)	(2)	(3)	(4)	(5)	(6)
Delayed gratification (DG)	0.071	−0.017	0.081	0.030	0.083	0.077
	(0.097)	(0.110)	(0.093)	(0.105)	(0.094)	(0.144)
Math comp. below avg.	0.521***	0.412***	0.383***	0.319**	0.606***	0.601***
	(0.111)	(0.119)	(0.122)	(0.143)	(0.095)	(0.116)
DG × Math comp. below avg.		0.416*		0.244		0.013
		(0.217)		(0.211)		(0.208)
Perceptual speed std	0.102	0.095	0.048	0.043	0.021	0.021
	(0.064)	(0.064)	(0.067)	(0.066)	(0.058)	(0.060)
Reasoning std	−0.022	−0.020	0.068	0.068	0.128***	0.128***
	(0.047)	(0.047)	(0.058)	(0.058)	(0.046)	(0.047)
Months between tests	0.120**	0.116*	0.189***	0.185***	0.190***	0.190***
	(0.060)	(0.060)	(0.053)	(0.053)	(0.056)	(0.057)
East German	−0.091	−0.099	−0.201	−0.205	−0.131	−0.131
	(0.135)	(0.132)	(0.127)	(0.127)	(0.117)	(0.118)
Male child	0.075	0.099	0.081	0.095	0.083	0.084
	(0.095)	(0.093)	(0.110)	(0.111)	(0.105)	(0.107)
Age in months std	−0.074	−0.048	−0.279	−0.264	−0.122	−0.122
	(0.188)	(0.187)	(0.210)	(0.208)	(0.187)	(0.188)
Extraversion std	−0.012	−0.004	0.048	0.053	0.044	0.044
	(0.062)	(0.061)	(0.062)	(0.063)	(0.057)	(0.057)
Conscientiousness std	−0.049	−0.051	0.026	0.024	0.054	0.054
	(0.059)	(0.060)	(0.059)	(0.059)	(0.052)	(0.052)
Agreeableness std	0.033	0.033	−0.026	−0.026	−0.039	−0.039
	(0.056)	(0.057)	(0.049)	(0.049)	(0.050)	(0.050)
Openness/Intellect std	−0.091	−0.091	0.020	0.020	0.133*	0.133*
	(0.082)	(0.082)	(0.062)	(0.061)	(0.068)	(0.068)
Neuroticism std	−0.045	−0.047	0.109**	0.108**	0.127**	0.127**
	(0.053)	(0.053)	(0.053)	(0.053)	(0.050)	(0.050)
Migration background	0.074	0.032	0.040	0.015	−0.153	−0.153
	(0.218)	(0.208)	(0.206)	(0.200)	(0.146)	(0.147)
Intermediate sec. educ.	0.595**	0.601**	0.072	0.077	−0.248	−0.248
	(0.283)	(0.276)	(0.303)	(0.302)	(0.257)	(0.257)
Univ. entrance qualif. or more	0.711**	0.713**	0.221	0.222	−0.076	−0.076
	(0.300)	(0.295)	(0.297)	(0.297)	(0.257)	(0.259)
More than 100 books at home	0.081	0.076	0.099	0.097	0.032	0.032
	(0.112)	(0.112)	(0.116)	(0.116)	(0.101)	(0.101)
Household income	−0.040	−0.028	0.005	0.012	0.029	0.029
	(0.080)	(0.079)	(0.074)	(0.075)	(0.055)	(0.055)
Living together with a partner	−0.120	−0.109	−0.023	−0.016	0.116	0.116
	(0.220)	(0.222)	(0.180)	(0.180)	(0.209)	(0.207)
Number of siblings	−0.024	−0.028	0.008	0.006	0.016	0.016
	(0.053)	(0.053)	(0.059)	(0.059)	(0.061)	(0.061)
Constant	−1.018	−0.922	−2.654**	−2.541**	−1.472**	−1.467**
	(0.813)	(0.815)	(1.120)	(1.122)	(0.602)	(0.609)
N	370	370	370	370	370	370
adj. R^2	0.049	0.054	0.059	0.059	0.147	0.144
Wald test: p-value[a]		0.036		0.143		0.501

Notes: Data: NEPS SC2 5.1.0, own calculations. Standard errors, reported in parentheses, are clustered at kindergarten group level.
[a] Test of hypothesis H_0: Delayed gratification (DG) + DG × Math comp. below avg. = 0.
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A.6: Different standard error calculations

	Coefficient	Level of standard error clustering		
		none	kindergarten group	classroom
		Panel A: Effects on competence level		
Kindergarten	0.235	0.109**	0.101**	
Grade 1	0.224	0.125*	0.122*	0.123*
Grade 2	0.232	0.127*	0.106**	0.113**
		Panel B: Effects on competence development		
Kindergarten - Grade 1	0.009	0.119	0.119	0.116
Kindergarten - Grade 2	0.038	0.124	0.110	0.122
Grade1 - Grade 2	0.022	0.111	0.123	0.115

Notes: Data: NEPS SC2 5.1.0, own calculations. The left column displays the respective coefficient for the decision to delay in the main estimations including general cognitive skill measures (tables 2.2 & 2.3, columns (2),(4) and (6)). The three columns on the right display the respective standard errors produced by different levels of standard error clustering.
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A.7: Competence development from Grade 1 to Grade 2 (Control variables from Grade 1. Standard errors, reported in parantheses, are clustered at classroom level.)

	(1)	(2)	(3)	(4)	(5)	(6)
Delayed gratification (DG)	−0.026	0.052	0.120	−0.038	0.049	0.119
	(0.132)	(0.125)	(0.201)	(0.130)	(0.122)	(0.197)
Math comp. below avg.		0.550***	0.611***		0.646***	0.712***
		(0.118)	(0.140)		(0.126)	(0.156)
DG × Math comp. below avg.			−0.167			−0.173
			(0.280)			(0.282)
Perceptual speed std				−0.062	0.036	0.042
				(0.075)	(0.074)	(0.077)
Reasoning std				0.127*	0.168***	0.166***
				(0.065)	(0.063)	(0.063)
N	221	221	221	221	221	221
adj. R^2	0.034	0.102	0.100	0.043	0.129	0.126
Wald test: p-value[a]			0.769			0.746

Notes: Data: NEPS SC2 5.1.0, own calculations. All estimations contain a constant, control for months between tests and all other explanatory variables named in table 2.1.* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.
[a] Test of the hypothesis H_0: Delayed gratification (DG) + DG × Math comp. below avg. = 0.

Table A.8: Effects on Math competence level (Extended version of table 2.2)

	Kindergarten			Grade 1			Grade 2		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Delayed gratification (DG)	0.317***	0.233***	0.276***	0.299***	0.217**	0.252**	0.313***	0.222**	0.249**
	(0.089)	(0.085)	(0.087)	(0.104)	(0.103)	(0.103)	(0.103)	(0.096)	(0.097)
Perceptual speed std		0.243***	0.237***		0.293***	0.281***		0.218***	0.217***
		(0.048)	(0.050)		(0.056)	(0.053)		(0.063)	(0.067)
Reasoning std		0.254***	0.266***		0.205***	0.207***		0.310***	0.325***
		(0.044)	(0.045)		(0.051)	(0.050)		(0.060)	(0.061)
Personality measures	Yes	Yes	No	Yes	Yes	No	Yes	Yes	No
N	370	370	370	370	370	370	370	370	370
adj. R^2	0.199	0.331	0.312	0.131	0.229	0.226	0.153	0.271	0.262

Notes: Data: NEPS SC2 5.1.0, own calculations. All estimations contain a constant and control for all other explanatory variables named in table 2.1. A full table with all controls is also displayed in the appendix table A.3. Standard errors, reported in parentheses, are clustered on kindergarten group level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A.9: Math competence development (Extended version of table 2.3)

	Kindergarten - Grade 1			Kindergarten - Grade 2			Grade 1 - Grade 2		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Delayed gratification (DG)	−0.003	−0.003	−0.013	0.039	0.028	0.010	0.037	0.025	0.017
	(0.100)	(0.100)	(0.103)	(0.095)	(0.094)	(0.091)	(0.099)	(0.098)	(0.104)
Perceptual speed std		0.053	0.045		0.013	0.017		−0.037	−0.025
		(0.065)	(0.065)		(0.069)	(0.071)		(0.058)	(0.059)
Reasoning std		−0.043	−0.054		0.052	0.053		0.089*	0.101**
		(0.047)	(0.048)		(0.058)	(0.060)		(0.048)	(0.049)
Personality measures	Yes	Yes	No	Yes	Yes	No	Yes	Yes	No
N	370	370	370	370	370	370	370	370	370
adj. R^2	0.002	0.000	0.000	0.037	0.035	0.034	0.056	0.060	0.032

Notes: Data: NEPS SC2 5.1.0, own calculations. All estimations contain a constant, control for months between tests and all other explanatory variables named in table 2.1. Standard errors, reported in parentheses, are clustered at kindergarten group level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A.10: Effect heterogeneity by initial math competence (Extended version of table 2.4)

	Kindergarten - Grade 1			Kindergarten - Grade 2			Grade 1 - Grade 2		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Delayed gratification (DG)	0.071	−0.017	−0.015	0.081	0.030	0.017	0.083	0.077	0.074
	(0.097)	(0.110)	(0.112)	(0.093)	(0.105)	(0.104)	(0.094)	(0.144)	(0.150)
Math comp. below avg.	0.521***	0.412***	0.417***	0.383***	0.319**	0.334**	0.606***	0.601***	0.615***
	(0.111)	(0.119)	(0.121)	(0.122)	(0.143)	(0.140)	(0.095)	(0.116)	(0.117)
DG × Math comp. below avg.		0.416*	0.399*		0.244	0.259		0.013	0.024
		(0.217)	(0.226)		(0.211)	(0.204)		(0.208)	(0.203)
Personality measures	Yes	Yes	No	Yes	Yes	No	Yes	Yes	No
N	370	370	370	370	370	370	370	370	370
adj. R^2	0.049	0.054	0.054	0.059	0.059	0.061	0.147	0.144	0.122
Wald test: p-value[a]		0.036	0.057		0.143	0.119		0.501	0.449

Notes: Data: NEPS SC2 5.1.0, own calculations. All estimations contain a constant, control for months between tests and all other explanatory variables named in table 2.1. Standard errors, reported in parentheses, are clustered at kindergarten group level * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.
[a] Test of the hypothesis H_0: Delayed gratification (DG) + DG × Math comp. below avg. = 0.

Chapter 3

Does the Transition into Daylight Saving Time Affect Students' Performance?

Stefanie P. Herber, Johanna Sophie Quis, and Guido Heineck

This study has been published in *Economics of Education Review* 61 (2017), 130-139:

https://doi.org/10.1016/j.econedurev.2017.07.002

Chapter 4

Does higher learning intensity affect student well-being? Evidence from the National Educational Panel Study

Johanna Sophie Quis

A version of this study containing only health outcomes has been published in *Journal of Economics and Statistics* 238(5), (2018), 441–476: https://doi.org/10.1515/jbnst-2018-0004

4.1 Introduction

Starting in 2000, most German *Länder* reduced secondary school duration in the academic track (*Gymnasium*) from nine (G9) to eight (G8) years, thereby responding to the comparatively high age of German academic track graduates and unsatisfactory results from the first round of PISA tests. Importantly, while school duration was reduced from nine to eight years, the federal requirements for cumulative instruction time for high school students remained the same. Therefore, the students' learning intensity increased substantially. The following scientific debate addressed whether this increase affected students' skills and academic performance (e. g. Büttner and Thomsen, 2015; Thiel et al., 2014). Psycho-social outcomes of the G8-introduction were widely discussed in the public, but rarely addressed in the scientific literature. The public debate about G8 was mostly shaped by parents fearing an overwhelming workload for their children. Accordingly, a variety of newspaper articles stated that students' stress levels increased dramatically and that the *"Turbo-Abitur"* was preventing children from having a normal childhood with enough time to play, recreate, and develop personally (e. g. Sussebach, 2011; Karakurt, 2011).[1] The current chapter investigates the effects of the shorter school duration imposed by G8 on students' psycho-social outcomes such as stress, mental health and well-being.

A recent and growing body of literature provides evidence that adolescents' psycho-social status matters for long-term socioeconomic outcomes, like unemployment status, or income. For example, male adolescents with mental health conditions like neurosis or personality disorders are more often unemployed, receive more social benefits and receive lower income throughout

[1] Conversely, a smaller number of newspaper articles claimed that the Abitur after 12 years has been common in Saxony and Thuringia, causing no such problems and that reports of student overload were single cases exaggerating a rather small problem that exists only for low performing academic track students. For further details on the entirety of the debate about G8 please refer to the comprehensive summary by Kühn et al. (2013) as well as Homuth (2017, p. 85f), who discusses the public debate in more detail.

adulthood (Lundborg et al., 2014). Smith and Smith (2010) report, that adults with psychological problems during childhood earn substantially less than adults without problematic childhoods.

In this chapter, I focus on internalizing mental health problems as measure for mental health. Symptoms of internalizing mental health problems typically manifest within the individual, examples are social retraction, physical (psychosomatic) conditions, anxiety and depression (Bilz, 2008, p. 45). Carneiro et al. (2007) observe that depression at age 11 reduces the probability of attaining educational degrees, increases the probability of heavy smoking at age 16, being expelled from school, and reporting symptoms of depression at age 42. Reivich et al. (2013) point out that depression and also already sub-clinical symptoms of depression in childhood and adolescence are stable over the life course and interfere severely with the child's ability to function. Taken together, this suggests that experiencing internalizing mental health problems as child reduces the psychological and socioeconomic well-being as adult. Therefore, the question arises whether the G8 reforms increased such internalizing mental health problems.

One cause of internalizing mental health problems in children lies in school-related issues: Crystal et al. (1994) report that for adolescents in grade 11 from the United States, Japan and Taiwan, school-related problems are the main source of stress, and that school and peers are the most frequently stated reasons for depressive feelings. Accordingly, the situation in school affects perceived stress and to some extent also mental health problems in several cultures.

Concerning student well-being, Lévy-Garboua et al. (2006), for example, demonstrate that adolescents reporting unhappiness and dissatisfaction are more likely to engage in risky or undesirable behaviors (among others: smoking, drug use, school absenteeism and violence). De Neve and Oswald (2012) conclude that adolescents who report higher life satisfaction and positive affect have higher income in later life. Mediating pathways of these findings are higher probabilities of obtaining a college degree, being optimistic and extroverted, less neurotic, and being hired and promoted more often.

To date, only few studies examine the impact of the German G8 school reform on high school students' psycho-social situation: Quis and Reif (2017) and Hofmann and Mühlenweg (2017) examined the effects of G8 on students' health, both when they were still in school and after graduation with mixed results. Milde-Busch et al. (2010) found that there is no higher prevalence of stress-related symptoms like headache among G8-students in a sample of students in grade 10 and 11. In contrast to their study, I observe students at graduation, where they should have attained the same level of skills and am using indexes of internalizing mental health problems and perceived stress instead of presenting a long list of symptoms. I also broaden the outcomes to well-being in school and life in general. Additionally, I provide several robustness checks to increase the credibility of the results.

International evidence on shortened high-school duration is also rare. Shorter school duration decreased performance in Canada (Morin, 2015; Krashinsky, 2014) but left mathematical performance unaffected in Switzerland (Skirbekk, 2006). Meanwhile, the evidence on how more school days per year affect performance is mixed. While Lee and Barro (2001) report no effects of increased instruction time, others find significantly positive effects on learning (Andersen et al., 2016) and test performance (Lavy, 2015; Eide and Showalter, 1998).

The current study answers two research questions: The first question asks, whether the reduction in school duration and the increase in schooling intensity caused by the G8 reform affected several measures of student well-being. The second question asks, which types of students are more likely to be affected by the reform. In particular, I pay special attention to the differences between male and female students, because both occurrence and sensitivity to mental health problems in adolescence are substantially higher for females (Bilz, 2008; Forehand et al., 1991; Crystal et al., 1994; Nolen-Hoeksema and Girgus, 1994).

I use data from the *Additional Study Baden-Wurttemberg* of the National Educational Panel Study (NEPS), which was conducted to evaluate the introduction of G8. The results indicate that the implementation of G8 for all academic track students increased females' perceived stress and symptoms of internalizing mental health problems, while for males only stress increased and mental health was not affected. However, the effects of G8 on stress and internalizing mental health problems do not carry over to broader measures of well-being in school and life in general to sustain a more general notion of well-being. The observed effects might either be caused by implementation problems of the reform and specific to the double cohort observed in the data or a general consequence of the implementation of G8 or a combination of both. My results imply that shortening school duration without shortening the teaching load can have a detrimental effect on a large number of students' mental health and stress levels. This should be considered for future reforms of school systems.

4.2 The G8 reform and debate

This section briefly introduces the German school system, then describes the G8 reforms and finishes with an overview of the scientific debate that was induced by the reform. For a more thorough explanation of the German school system refer to Lohmar and Eckhardt (2013).

4.2.1 The German school system

The German school system is not completely uniform across federal states because education policies lie within responsibility of the 16 federal states. A simplified version is displayed in figure 4.1. Commonly, students enter primary school at the age of 6 years. In most states, all students attend primary school for four years.[2] In secondary school, students are usually separated into three different tracks. The basic and intermediate tracks, *Hauptschule* and *Realschule*, qualify students for vocational training. The academic track, *Gymnasium*, usually lasts from grade 5 to 12 or 13, depending on whether the student is enrolled in a G8 or G9 track. Successful completion of *Gymnasium* leads to the *Abitur*, the formal entry qualification for higher education in Germany.

While education policies are determined at state level, the *Kultusministerkonferenz*[3] sets binding requirements for the educational framework at the federal level to ensure comparability of schools and degrees within Germany.

4.2.2 Reforms shortening school duration

The G8-reforms reduced secondary school duration from nine to eight years in most German states after 2000.[4] The reforms targeted several aims: One was to speed up the job market entry of high-school graduates to increase the international competitiveness of German graduates from secondary and tertiary academic track education (Ministerium für Bildung Kultur und Wissenschaft, Saarland, 2000). Graduates of the academic track schools were and are, in international comparison, old when they enter the labor market, because they spent relatively long time in

[2] In some three states primary school encompasses six years.

[3] The Standing Conference of the Ministers of Education and Cultural Affairs of the Länder in the Federal Republic of Germany.

[4] There has been a history of G8 before 2000: Prior to the German reunification in 1990, all East German states had school systems leading to the *Abitur* after 12 years, while all West German states required 13 years. After 1990, most East German states adopted the West German G9 system, while Saxony and Thuringia maintained the G8 system. Additionally, some states, including Baden-Wurttemberg already offered G8 classes in selected schools as fast-track programs to particularly skilled students before G8 was implemented as a general system.

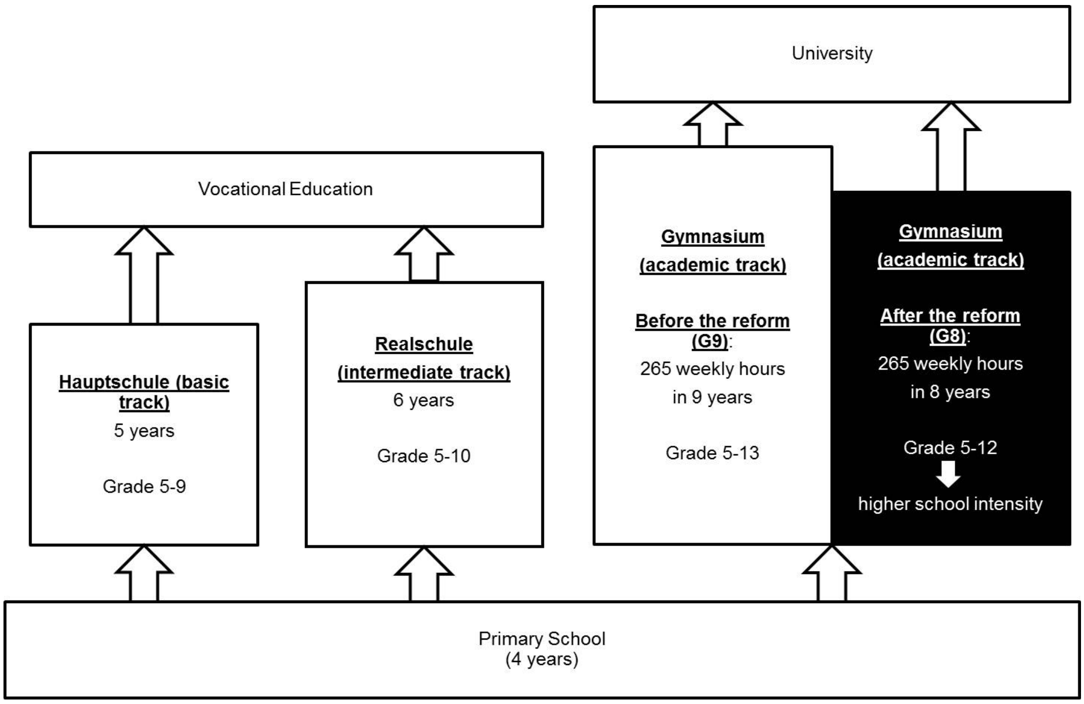


Figure 4.1: The German School System

secondary school[5] and often afterwards longer time in university[6] than on international average (Büttner and Thomsen, 2015). In most OECD states primary and secondary schooling have a combined duration of 12 years, whereas it lasted 13 years in Germany.

A second aim was to lengthen work-lives in order to reduce the burden of an aging population on the social security systems not only by increasing retirement age but also by a reduction of age at the labor market entry (Büttner and Thomsen, 2015). The third reasoning behind the G8 reforms concerned student performance. The overall German performance in the first round of PISA tests in 2000 was disappointing, students from the two federal states with a G8 system at that time – Thuringia and Saxony – performed above the German PISA average. This additionally increased the demand for shortening school duration. Further, some claimed the reform would reduce the costs of schooling (Wiater, 1996), and increase schooling efficiency (Schavan and Ahnen, 2001).

In contrast, there were also concerns that G8 might reduce time for extra-curricular activities, and impede personal and psycho-social development (Wiater, 1997). Further, cost-reductions through reduced school duration could be outweighed by additional costs through curriculum revisions in the short term and lengthening schooldays into the afternoons in the long run. Finally, increasing students' time pressure could lead to superficial learning and reduced performance levels of students (Dam, 2005; Altner, 2007; Brütting, 2007).

Importantly, the G8 reforms (reduction in school duration), left the joint requirements for graduation at the federal level unaffected (Kultusministerkonferenz, 2013). This means that stu-

[5] The typical graduation age for the German Abitur was 19 in 1998 (OECD, 2000, p. 319), and with 19 to 20 years even higher in 2011 (OECD, 2013a, p. 408). In other OECD countries like the United States, Japan, or France it was between 17 and 18 (OECD, 2013a, p. 408).

[6] Their graduation age was 25 to 26 years in 1998, while at that time it was 21 to 24 in the United States, Japan and France (OECD, 2000, p. 320), and 25 to 27 years in 2011 (OECD, 2013a).

dents had to complete 265 *Jahreswochenstunden*[7] between grade 5 and graduation, irrespective of whether they attended a G8 or G9 track. Concerning academic requirements, the unified examination requirements also remained in place. Compared to G9 students, G8 students were therefore exposed to a substantially increased schooling intensity.

The introduction of G8 started in 2001 in Saarland where the first cohort of G8 students graduated in 2009. Nine states introduced G8 between 2002 and 2005, three states followed until 2008. Baden-Wurttemberg introduced G8 in 2004/05, with the first G8-cohort graduating in 2012 together with the last G9 cohort.

In contrast to most other states, Baden-Wurttemberg revised all curricula and moved from an input-oriented teaching plan to an output-oriented education plan, thereby reducing the amount of topics to be covered in some subjects. For the students from Baden-Wurttemberg, this implied different curricula from grade 5 through 11 (G9) and 5 through 10 (G8). However, in their final two years, i. e. in grade 12-13 for G9- and in grade 11-12 for G8-students, the first G8-students and the last G9-students were instructed together in the same classroom and faced the same requirements in class as well as in the final examinations. As the last G9 cohort and the first G8 cohort graduated in the same year, they are referred to as *double cohort*. To avoid complications, the curriculum for the final two years of the double cohort was specifically designed to be based only on the junction between the curricula for G8 and G9. This means, again, that in comparison to the G9-students, G8-students had to acquire the same amount of skills in a shorter period of time, thereby increasing the learning intensity. An increase in learning intensity directly implies shorter education duration and therefore leads to graduation at a younger age. I consider the resulting age difference at graduation as part of increasing learning intensity.[8]

Table 4.1: Anticipated effects of the reform

Outcome	**Effect**	**Suspected mechanism**
Stress	+	Increase in weekly instructional time
		Less time for other activities
		Higher demands in class
		Potentially more competition
Mental health problems	+	High levels of stress increase mental health problems
Satisfaction with school	−	Higher demands for affected students
		More stressors
	+	More time with friends in school
		School duration decreases by one year
General well-being	+ − 0	dependent on perceived stress and well-being in school

Due to the reform, in a normal school week, students spent more time in school to attain the required cumulative instruction time and potentially still had to spend the same amount of time on homework, tutoring, exercise and preparation as G9 students, to acquire the given material within the shortened time period.[9] The reduction in school duration led to a higher learning speed in class, which might have been an additional burden for some students. Further, the competition between students from G8 and G9 might be pronounced especially in the double cohort. Students in school, who can already foresee increased competition for scarce resources directly after graduation (e. g. apprenticeships, volunteering activities, university places, student accommodation in university cities,…), might additionally suffer from the reforms.

[7] This is a cumulative measure of weekly instruction hours over several school years. For example, 29 weekly instruction hours in grade 5 and 30 weekly instruction hours in grade 6 sum up to 59 *Jahreswochenstunden*.

[8] However, I also discuss the implications of age differences in chapter 4.6.

[9] 265 weekly instruction hours within nine years means, on average, 29.44 weekly instruction hours per year, while the same cumulative instruction time requires an average of 33.1 weekly instruction hours if spread over eight years.

When nothing else changes, especially not the coping mechanisms of students, I expect perceived stress levels to rise. Since an increase in stress is related to an increase in psychic and psychosomatic problems (Bergmüller, 2007, p. 166), I expect the reform to negatively influence mental health of the students, i. e. increase their internalizing mental health problems.

On the other hand, more time spent in school might also mean, students spend more time with their friends in school. Therefore, the effect of increased stress on student general life satisfaction might be reduced by an increase in satisfaction with the social aspects of school life. Since, according to van Praag and Ferrer-i Carbonell (2010), general subjective well-being can be seen as a linear composition of well-being in several domains, one effect might be able to outweigh another. Additionally, a perceived pressure from outside, e. g. through media coverage or parents and teachers, might also lead to a stronger spirit of solidarity among students which makes ex ante predictions of the reform effect on general subjective well-being and on satisfaction with the situation in school difficult. Therefore, predictions of the reform effect on well-being in terms of satisfaction with life and school remain ambiguous.

4.2.3 Prior findings on the effects of the G8 reform

In the interdisciplinary literature concerning G8 three types of studies prevail. The first type concerns fast-track classrooms, programs where the best students were sorted into a separate classroom in order to graduate one year earlier. Several studies positively evaluated these programs (Kaiser and Kaiser, 1998; Heller, 2002; Zydatiß, 1999). One limitation of these findings is that the results regarding highly-skilled student populations cannot readily be extrapolated to the general student population affected by the G8 reforms.

The second type consists of between-state comparisons of school systems. Böhm-Kasper and Weishaupt (2002) evaluated G8 as general system in a comparison of two G9 states with one G8 state finding no differences regarding students' burdens attributable to G8/G9. A shortcoming of this study is that the federal states are responsible for schooling policies. Therefore, a direct comparison between states might capture differences in policies between states other than G8/G9-differences or G8/G9-differences might be attributed to inter-state differences.

The third type of studies evaluates the introduction of G8 as a general system. This work either uses within state comparisons of the situation before and after the introduction of G8, or a difference-in-differences framework over several states. The current chapter follows the former approach. Most studies investigate outcomes that are not related to well-being in a broader sense. The introduction of G8 led to lower grades at graduation (Büttner and Thomsen, 2015; Huebener and Marcus, 2017), while the impacts on student skills were mixed (Dahmann, in press; Huebener et al., in press; Hübner et al., 2017), and there were only marginal effects on personality traits (Thiel et al., 2014; Dahmann and Anger, 2014). There seems to be no impact on performance in university (Dörsam and Lauber, 2015; Kühn, 2014; Meyer and Thomsen, 2017), which could be due to selection into university since university enrollment was delayed or abandoned more often after the introduction of G8 (Meyer and Thomsen, 2016; Meyer et al., forthcoming; Marcus and Zambre, forthcoming). Concerning the previously stated goal of an earlier job market entry, Huebener and Marcus (2015), using administrative data, attribute a reduction of graduation age by 10 months to the reform. At least part of the difference to a full year, which could have been expected, is linked to a sharp increase in G8-students' grade repetition during the penultimate year of schooling. The Autorengruppe Bildungsberichterstattung (2010) reports less mobility from lower tracks into the academic track after the reform. For a more extensive discussion refer to the literature review by Thomsen (2015).

The evidence regarding psychological effects of the G8 reform is mixed. Using a sample of students from two cities in Saxony-Anhalt, Meyer and Thomsen (2015) found that students reported more burdens from learning, while they spent less time on side jobs and volunteering after the introduction of G8. Westermaier (2016) evaluated the impact of the reform on crime rates and substance abuse. He finds a lower crime rate which is driven by less violence and less substance abuse, especially decreased cannabis consumption, for G8 students. Milde-Busch et al. (2010) descriptively compared G8-students in grade 10 and G9-students in grade 11 from Munich, but failed to find substantial differences between G8 and G9 students' reported headaches and similar symptoms. However, G8-students reported less leisure time to compensate for school stress. Meyer and Thomsen (2017) found health differences between university students who were subject to G8 and those who were subject to G9: Females reported better health and males reported worse health when they were subject to G8.

Given that students who experience more psychological problems during school will likely be socioeconomically disadvantaged as adults, it is important to study if G8 increased mental health problems of students. If psychological problems increase after the introduction of G8, policy makers and schools should consider measures to target the psychological well-being of G8 students if they want to unfold the full socioeconomic potential of adults that have undergone G8.

To sum up, there is a heated public debate especially about the burdens of G8 as compared to G9. Nevertheless, there is only little empirical knowledge about the actual effects of G8 on well-being. In my analysis, I contribute to closing the research gap with respect to subjective stress, internalizing mental health problems and general well-being in a larger within-state sample.

4.3 Data

The data used for this analysis are drawn from the second wave of the Additional Study Baden-Wurttemberg of the National Educational Panel Study (NEPS).[10] The purpose of the Additional Study Baden-Wurttemberg was to answer whether the shortening of school duration by one year affected student achievement, cognitive functioning, well-being, leisure-time activities, and participation in private tutoring (Wagner et al., 2011, p. 245). The data consist of three waves, which encompass the last pure G9 cohort (2010/11), the double cohort (2011/12), and the first pure G8 cohort (2012/13). Each wave is a cross-section of students in their final school year (grade 12 for G8 students, grade 13 for G9-students), therefore, each student's data was collected only once. The study included a survey and competence tests for students; additional surveys were completed by teachers and school principals. The testing took place during school lessons after the *Abitur* examinations were written, but before their results were announced. As an additional incentive for students to take part in the survey each participant was paid EUR 10 in cash after completion of the test. The second wave was sampled in late spring 2012 in 48 academic track schools.

Table 4.2 summarizes the analysis sample. A total of 2391 students in their final school year in 2011/2012 in Baden-Wurttemberg were sampled. In total, 32 observations were dropped due to missing values on dependent variables,[11] 53 observations were dropped due to missing information on explanatory variables.[12] Therefore, the analysis sample consists of 2306 observations.

Since Baden-Wurttemberg offered a fast-track program for particularly high performing students in some schools, the first wave (2010/11) also includes 52 students who were already subject to G8 prior to the general introduction of G8. The data do not allow to distinguish the fast-track

[10] This paper uses data from the National Educational Panel Study (NEPS): G8 Reform Study in Baden-Wurttemberg, doi 10.5157/NEPS:BW:3.0.0. The NEPS data collection is part of the Framework Programme for the Promotion of Empirical Educational Research, funded by the German Federal Ministry of Education and Research and supported by the Federal States. (Blossfeld et al., 2011).

[11] For students with more than two missing items within the stress scale or more than four missing items on the internalizing mental health problem scale, the respective index was set to missing.

[12] Definitions of all variables are given in table A4.1.

students from regular G8 students in the second wave (2011/12), a problem that is further discussed in the robustness checks (chapter 4.6).

In my analyses, indexes of perceived stress and common symptoms of internalizing mental health problems (e. g. anger, sadness, exhaustion, and headaches), as well as measures of subjective well-being (self-assessed life and school satisfaction) serve as outcomes. The question for stress was developed by NEPS and contains 15 items, the question for internalizing mental health problems originates from the PISA 2003 questionnaires and contains 26 items. I used the description on the construction of the second index by Bergmüller (2007) to build both indexes for this study. Tables A4.2 and A4.3 in the appendix provide a detailed list of the items contained in the indexes as well as the response formats in detail. Both indexes are on a scale between 1 and 4, where 4 displays highest stress levels ("exactly true") and strongest internalizing mental health problems (symptoms occurred more than six times in the last six weeks).[13]

In the descriptive statistics (table 4.2) the indexes are reported as average scores. The stress score is the average answer students have given on all 15 stress items. In the sample the average stress score is 2.1; that means on average students answered all items with "hardly true". An increase (decrease) by one standard deviation (.59), translated into answers on the items, would mean that 8.9 of the 15 items were answered with a one point higher (lower) rating.

The average mental health score is 1.9 (on average most symptoms occurred once or twice in the last six weeks), with a standard deviation of .55. As 26 items count into the score, a one standard deviation increase in the score amounts to 14.3 items that increased by one point (i. e. from once or twice to three to six times in the last six weeks). In order to simplify the interpretation of the regression results, the average scores were transformed into indexes by standardizing the average scores. All impacts on stress and internalizing mental health problems can therefore be interpreted as fractions of one standard deviation.

Well-being in life and well-being in school were both measured on an 11-point Lickert-type scale where 10 represents the highest level of well-being. The interpretation of results on the well-being measures is therefore straightforward.

The set of background variables provides information on books at home, age, previous grade repetitions, sports participation,[14] and migration background. All of these and, in addition, students' gender form important control variables because they are likely to influence student well-being themselves.

The analysis sample contains 43% males and 57% females. This is a slightly higher share of females compared to 53% in higher secondary school in Baden-Wurttemberg as reported by the German federal statistical office (Statistisches Bundesamt, 2013, table 3.1). One fifth of the sampled students have a migration background in the sense that at least one of their parents was not born in Germany. About 5% repeated at least one grade. The sampled students in G9 are on average more than one year older than the G8 students. This might be attributed to the large share of grade repeaters among G9 students (9%) as compared to G8 students (1%). Because for students from the last G9-cohort repeating a grade and therefore ending up in the first G8-cohort graduating at the same point in time would not have made sense, this difference is inherent in the nature of the reform.[15]

---

13 In chapter 4.6 I discuss how the results are affected by different calculations of the indexes.

14 Since one could consider sports participation also as an outcome of the reform, I also estimated without controlling for sports participation in the robustness section. This does not affect the resulting coefficients.

15 Tables A4.4 - A4.6 show how all variables change between the three waves of the Additional Study Baden-Wurttemberg. The numbers for previous grade repetition are substantially lower among G8-students in the double cohort but return to their old (pre-G8) levels in the first pure G8 cohort.

Table 4.2: Descriptive statistics by reform and by gender

		By reform		By gender		Pr $> \mid t \mid$	
	Pooled (1)	G9 (2)	G8 (3)	Male (4)	Female (5)	(2)=(3) (6)	(4)=(5) (7)
Outcomes							
Stress score	2.10	2.01	2.19	1.96	2.21	0.000***	0.000***
	(0.59)	(0.56)	(0.60)	(0.55)	(0.58)		
Mental health problem score	1.90	1.87	1.92	1.71	2.03	0.030**	0.000***
	(0.55)	(0.54)	(0.55)	(0.48)	(0.56)		
Well-being school	7.07	7.13	7.00	7.16	7.00	0.130	0.066*
	(2.03)	(2.05)	(2.02)	(2.05)	(2.02)		
Well-being life	7.74	7.70	7.78	7.82	7.68	0.246	0.047**
	(1.66)	(1.76)	(1.55)	(1.61)	(1.68)		
Student characteristics							
Reform (G8)	0.50	0.00	1.00	0.50	0.51		0.764
	(0.50)	(0.00)	(0.00)	(0.50)	(0.50)		
Female	0.57	0.57	0.58	0.00	1.00	0.764	
	(0.49)	(0.50)	(0.49)	(0.00)	(0.00)		
Age (years)	17.86	18.40	17.34	17.89	17.84	0.000***	0.078*
	(0.69)	(0.49)	(0.39)	(0.71)	(0.67)		
Repeated a grade	0.05	0.08	0.01	0.05	0.05	0.000***	0.308
	(0.22)	(0.28)	(0.12)	(0.23)	(0.21)		
Sports	0.77	0.78	0.76	0.84	0.72	0.394	0.000***
	(0.42)	(0.42)	(0.43)	(0.37)	(0.45)		
Parental background							
Migration backgr.	0.22	0.22	0.21	0.21	0.22	0.373	0.506
	(0.41)	(0.42)	(0.41)	(0.41)	(0.41)		
0-100 books at home %	0.19	0.18	0.19	0.20	0.17	0.496	0.056*
	(0.39)	(0.39)	(0.39)	(0.40)	(0.38)		
101-200 books at home %	0.17	0.16	0.17	0.16	0.17	0.579	0.507
	(0.37)	(0.37)	(0.38)	(0.37)	(0.38)		
201-500 books at home %	0.31	0.31	0.31	0.28	0.33	0.975	0.035**
	(0.46)	(0.46)	(0.46)	(0.45)	(0.47)		
500 + books at home %	0.34	0.35	0.33	0.35	0.33	0.332	0.313
	(0.47)	(0.48)	(0.47)	(0.48)	(0.47)		
Observations	2306	1147	1159	986	1320		

Notes: Data: NEPS BW D_3-0-0, wave 2011/12. Own calculations. Columns (1)-(5) display sample means and standard deviations in parentheses. Columns (6) - (7) display p-values from standard t-tests. Stress score is the mean of 15 symptoms (reported in table A4.2) on a range from 1 (not at all true) to 4 (exactly true). Mental health problem score is the mean of 26 symptoms (reported in table A4.3) on a scale from 1 (never occurred in the last 6 weeks) to 4 (symptom occurred more than 6 times in the last 6 weeks). Migration background is a dummy indicating whether at least one parent was not born in Germany.

4.4 Empirical strategy

Assignment to the G8 or G9 group can be assumed to be random conditional on the policy reform. It is not very likely that students skipped a year to stay in G9 or repeated a grade to study under the G8-regime. In both cases they would end up with the same amount of school years as they would have faced when obeying the initial assignment. Indeed, the share of grade repeaters is very low in the double cohort and returned to pre-reform levels in 2012/13. One option for students would be to move to another German Land which has not yet implemented the reform. As young students would likely need their parents to move along, this would be very costly. The other possibility would be to switch to a comprehensive school, but these were rare in Baden-Wurttemberg, therefore this option would only have been feasible for few students, as well. I therefore consider the introduction of G8 in Baden-Wurttemberg as natural experiment (Wagner et al., 2011, p. 243). Under this assumption, assignment to any group is independent of all factors determining the outcomes. Including a dummy that indicates whether the student was affected by the G8 reform will then pick up the causal reform effect. As the data used in this study is cross-sectional, I provide estimates of the effects of the reform by simple ordinary least squares estimation.

The econometric setup therefore has the following form:

$$Y_{is} = \alpha + G8_{is}\gamma + \mathbf{X}_{is}\mathbf{fi}_1 + \epsilon_{is}, \tag{4.1}$$

where Y_{is} is the respective outcome of student i in school s. $G8_{is}$ represents the assignment of student i in school s to the treatment (G8) or control (G9) group. $\mathbf{X}_{is}$ is a vector of student level variables measuring student and family background. The error term ϵ_{is} with mean zero is clustered at school level. Previous research has shown that adolescent females suffer from stress and internalizing mental health issues more strongly than males (e. g. Bergmüller, 2007; Bilz, 2008). The last column in table 4.2 shows that there are substantial gender differences in the present data. Therefore, I additionally conduct analyses separately for male and female students.

The controls used in this analysis include age relative to own cohort,[16] previous grade repetition,[17] sports participation,migration background, and as a proxy for socioeconomic background the amount of books at home. Even though age-deviation from the cohort median and grade repetition are correlated with each other, variance-inflation measures do not hint towards a collinearity problem, when controlling for both in one model. For the observable characteristics there are no substantial differences between the treatment and control group (table 4.2, column 6). The only exceptions are the differences in age and grade repeaters, which are inherent in the nature of the reform, as mentioned above.

4.5 Results

4.5.1 Perceived stress and internalizing mental health problems

The results for the standardized measure of perceived stress are reported in the first part of table 4.3. The reform increased stress levels in all samples. In the pooled sample, the reform effect amounts to 31% of a standard deviation, i. e. the average score increased by 0.18 (2.8 items increased by one category). The size of the effect seems to be driven by females who suffer more from the reform than males. In separate estimations, the treatment effect on the stress index is at 21% of a standard deviation (1.9 items) for males and 39% of a standard deviation (3.5 items

[16] The effects of the reform increase if age is included relative to both cohorts taken together (see chapter 4.6). It is therefore not likely, that otherwise nonexistent reform effects are driven by an age effect, I do not account for.

[17] In the robustness section I show that excluding these students does not alter the results from the main analyses.

Table 4.3: Regressions on stress and internalizing mental health problems

	Stress			**Mental health problems**		
	Pooled	Male	Female	Pooled	Male	Female
Reform (G8)	0.314***	0.214***	0.392***	0.111**	0.010	0.186***
	(0.049)	(0.065)	(0.063)	(0.048)	(0.061)	(0.063)
Age - median	0.015***	0.026***	0.006	0.003	0.004	0.001
	(0.005)	(0.007)	(0.006)	(0.005)	(0.007)	(0.006)
Repeated a grade	0.167	−0.054	0.365	0.334***	0.113	0.537***
	(0.127)	(0.168)	(0.220)	(0.123)	(0.150)	(0.172)
Female	0.424***			0.580***		
	(0.045)			(0.044)		
Sports	−0.234***	−0.199**	−0.257***	−0.128**	−0.161*	−0.110
	(0.051)	(0.087)	(0.057)	(0.049)	(0.082)	(0.068)
Migration backgr.	0.209***	0.242**	0.191**	0.241***	0.247***	0.241***
	(0.057)	(0.092)	(0.080)	(0.054)	(0.090)	(0.079)
Books at home:						
0-100 books	0.268***	0.294***	0.232***	0.123*	0.177**	0.068
	(0.055)	(0.079)	(0.085)	(0.064)	(0.085)	(0.090)
101-200 books	0.073	0.095	0.055	−0.067	−0.034	−0.101
	(0.053)	(0.072)	(0.067)	(0.062)	(0.072)	(0.083)
201-500 books	0.009	0.057	−0.026	−0.037	0.045	−0.100
	(0.046)	(0.056)	(0.072)	(0.052)	(0.059)	(0.073)
Constant	−0.345***	−0.349***	0.074	−0.360***	−0.311***	0.197**
	(0.059)	(0.083)	(0.075)	(0.065)	(0.091)	(0.074)
Observations	2306	986	1320	2306	986	1320
R^2	0.114	0.0695	0.0788	0.114	0.0317	0.0384

Notes: NEPS BW D_3-0-0 wave 2011/12. OLS regressions. The dependent variables are standardized to a mean of zero and a standard deviation of one. Age - median = age in months - median(age of students in own cohort). Books at home: reference category: 'more than 500'. A two-sided test for equality of the reform effect for the male and female sub-samples yields a p-value of 0.05 for stress and of 0.04 for internalizing mental health problems. Equality of the reform effect of males vs. females can therefore be rejected at the 5%-level.
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Standard errors, clustered at school level, in parentheses.

increased by 1) for females. A two-sided test for equality of the reform effect for males and females yields a p-value of 0.05. Equality of the effects can be rejected at the 5%-level. Even absent the reform, females report stress symptoms significantly more often. Hence, it seems that the reform aggravates previously existing gender differences in perceived stress. Furthermore, older male students report higher stress levels, while age relative to the cohort median has no effect for female students. Students with a migration background report more stress symptoms. Students who do sports at least twice weekly report lower levels of stress and the effect is stronger for females than for males. A low socioeconomic background, as proxied by amount of books at home, is associated with higher stress levels, even though not all effects are significant in the separated models.

The finding, that perceived stress increases more for females than for males is in line with other research showing that female adolescents tend to perceive stressors as more burdensome than males (Seiffge-Krenke, 2000; Compas and Wagner, 1991; Dornbusch et al., 1991). Females do not only report more stress, they also tend to respond to it with more internalizing symptoms than males, who in turn respond with more externalizing symptoms (Seiffge-Krenke, 2007; Grant et al., 2003). For German students, Bilz et al. (2003) show a strong positive relationship between perceived stress and psychosomatic problems for both sexes, with a stronger relationship for females. In line with this and a stronger increase in perceived stress for females, reform effects for females on the internalizing mental health problem index (table 4.3, second part) are, as expected, positive (i. e. internalizing mental health problems increased). While there seems to be no increase in mental health problems for males, the reform effect for the pooled sample amounts to 11% of a standard deviation, and is even higher for females (18% of a standard deviation i. e. 2.6 items increased by one). Age relative to own cohort has no significant explanatory power, the coefficient of grade repetition for females is sizable and significant, and sports participation is beneficial to mental health in the pooled and male sample, but not significant for females. Students with a migration background are over-proportionately affected by mental health problems. A low socioeconomic background, as measured by books at home, seems to matter only for male students' mental health.

Overall, the results imply a statistically significant increase in perceived stress and mental health problems through the G8 reform, which is mainly driven by females.

4.5.2 Subjective well-being in school and life in general

Whereas the findings for stress and mental health problems suggest large negative effects of the G8 reform particularly for females, satisfaction with the situation in school (table 4.4, first part) does not significantly change with the reform in the pooled and in the male sample. For females, the negative reform effect is marginally significant at the 10% level. In the pooled model, well-being in school is lower for students who are, compared to their classmates, older and in all models it is also lower for students with migration background or who have few books at home. Further, school satisfaction decreases substantially when a student has already repeated at least one grade, although the coefficients are only significant for the pooled sample and for females.

The decrease in female satisfaction with the situation in school does not affect general subjective well-being (second part of table 4.4). Students' satisfaction with life in general does not change significantly with the introduction of G8 in the pooled model. Furthermore, and unexpectedly, there is a positive, yet statistically insignificant, effect of the reform on male students' well-being. This result may be explained by the idea of general life satisfaction being a combination of several domain satisfactions (van Praag and Ferrer-i Carbonell, 2010, ch. 6). If satisfaction with the situation in class decreases, the decrease might be outweighed by another effect: More time spent in school also means for most students that they spend more time with their friends in school and hence are more satisfied with their friendships. Another possibility is, that male students dislike school in general and therefore leaving school one year earlier actually increases their well-being. Doing sports regularly increases well-being substantially, while previous grade repetition decreases it. There seems to be a negative, but statistically insignificant effect of a low socioeconomic background.

Overall, the data imply only a small negative effect of the reform on students' well-being in school and no effect on students' general subjective well-being. Since school satisfaction did not decrease a lot due to the reform, it seems plausible that general life satisfaction did also not decrease.

4.5.3 Effect heterogeneity

To test if the effects of G8 on perceived stress, internalizing mental health problems and well-being are driven or mitigated by subgroups - as already shown for gender - heterogeneity analyses with respect to age, student performance, books at home, and migration background were conducted. The results are displayed in table 4.5.[18]

[18] Results including separate estimations for males and females are provided in tables A4.7 & A4.8 in the appendix.

Table 4.4: Regressions on students' subjective well-being in school and life in general

	School			Life in general		
	Pooled	Male	Female	Pooled	Male	Female
Reform (G8)	−0.182	−0.082	−0.243*	0.065	0.143	0.002
	(0.112)	(0.155)	(0.130)	(0.070)	(0.095)	(0.093)
Age - median	−0.018*	−0.014	−0.020	0.004	−0.005	0.011
	(0.011)	(0.016)	(0.014)	(0.009)	(0.014)	(0.011)
Repeated a grade	−0.663***	−0.521	−0.761**	−0.323*	−0.076	−0.563*
	(0.231)	(0.326)	(0.356)	(0.189)	(0.270)	(0.288)
Female	−0.138			−0.088		
	(0.094)			(0.072)		
Sports	0.313***	0.428**	0.251*	0.418***	0.642***	0.312***
	(0.106)	(0.186)	(0.128)	(0.086)	(0.126)	(0.106)
Migration backgr.	−0.587***	−0.858***	−0.402***	−0.186**	−0.070	−0.267**
	(0.116)	(0.205)	(0.142)	(0.086)	(0.156)	(0.115)
Books at home:						
0-100 books	−0.501***	−0.435**	−0.535***	−0.135	−0.207	−0.079
	(0.139)	(0.203)	(0.192)	(0.095)	(0.159)	(0.138)
101-200 books	−0.323**	−0.177	−0.418**	0.012	0.016	0.012
	(0.130)	(0.186)	(0.182)	(0.105)	(0.175)	(0.146)
201-500 books	−0.004	−0.042	0.025	−0.044	−0.081	−0.017
	(0.089)	(0.141)	(0.135)	(0.074)	(0.138)	(0.120)
Constant	7.313***	7.187***	7.227***	7.527***	7.291***	7.555***
	(0.120)	(0.182)	(0.156)	(0.115)	(0.149)	(0.124)
Observations	2306	986	1320	2306	986	1320
R^2	0.0494	0.0568	0.0472	0.0192	0.0266	0.0169

Notes: NEPS BW D_3-0-0 wave 2011/12. OLS regressions. All models additionally contain a constant. Age - median = age in months - median(age of students in own cohort). Books at home: reference category: 'more than 500'. Dependent variable: 'How satisfied are you currently and in general terms, with your life?' and 'How how satisfied are you with your situation at school?' (11-point scales, translated by NEPS). * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Standard errors, clustered at school level, in parentheses.

Age– Potentially, the youngest students within a cohort lack the emotional and cognitive maturity to face the increased demands in a G8 system. Therefore results of the analyses including an interaction of the reform dummy with a dummy indicating whether the student was older than the median student in her cohort are displayed in table 4.5. For stress as well as both measures of well-being the results do not point towards effect heterogeneity by age. The p-value of the Wald tests shows that the effect of the reform on perceived stress and internalizing mental health problems for older students is significantly different from zero. The effects on mental health problems are slightly, yet not significantly higher for older students. This finding is in line with literature stating that the prevalence of mental health problems rises until late adolescence or early adulthood and then remains stable (Lewinsohn et al., 1998; Thapar et al., 2012; Patel et al., 2007; Merikangas et al., 2009, 2010).

Mathematical competence – Student ability could mediate the increase in psychological problems through G8 because highly skilled students might be less affected by the reform. A dummy indicating whether the student was above the median mathematical competence in her cohort was used as proxy for a high skill level. This dummy was interacted with the effect of the G8 reform.[19] Students with higher math competence show slightly, yet not statistically significantly lower levels of perceived stress than students with lower mathematical competence. However,

[19] Unfortunately, the data only provide current mathematical ability and not a pre-reform measure. If ability is indeed influenced by the reform (Dahmann, in press; Huebener et al., in press; Hübner et al., 2017), then maybe this is rather a mechanism than a real heterogeneity.

Table 4.5: Effect heterogeneity

	Perceived stress		**Mental health problems**		**Well-being school**		**Well-being life in general**	
Baseline model for reference								
Reform (G8)	0.314***	(0.049)	0.111**	(0.048)	−0.182	(0.112)	0.065	(0.070)
Age higher than median								
Reform (G8)	0.278***	(0.054)	0.056	(0.055)	−0.226*	(0.133)	0.028	(0.088)
Interaction	0.075	(0.062)	0.113	(0.068)	0.091	(0.135)	0.077	(0.118)
Wald test p-value[a]	0.000***		0.009***		0.299		0.273	
Math competence above median[b]								
Reform (G8)	0.373***	(0.050)	0.151**	(0.067)	−0.236*	(0.121)	0.015	(0.108)
Interaction	−0.118	(0.075)	−0.075	(0.085)	0.109	(0.170)	0.099	(0.146)
Wald test p-value[a]	0.001***		0.225		0.423		0.229	
More than 200 books at home (higher SES)								
Reform (G8)	0.265***	(0.064)	0.072	(0.079)	−0.185	(0.202)	0.066	(0.126)
Interaction	0.075	(0.076)	0.061	(0.091)	0.005	(0.209)	−0.001	(0.152)
Wald test p-value[a]	0.000***		0.022**		0.115		0.446	
Students with migration background								
Reform (G8)	0.341***	(0.057)	0.137***	(0.048)	−0.237*	(0.119)	0.055	(0.070)
Interaction	−0.127	(0.102)	−0.124	(0.095)	0.259	(0.248)	0.048	(0.196)
Wald test p-value[a]	0.016**		0.890		0.927		0.587	
Observations	2306		2306		2306		2306	

Notes: NEPS BW D_3-0-0 wave 2011/12. OLS regressions. All models contain a constant and control additionally for all variables stated in table 4.3 and a dummy indicating whether the student's mathematical competence is above the median in her cohort (only mathematical competence part). The interaction term displays the interaction of the reform with being older than the median student, with above median mathematical competence, with having more than 200 books at home or with having a migration background.
[a]Test of the hypothesis H_0: Reform (G8) + Interaction = 0.
[b]Due to missing data, 3 observations are missing in the mathematical competence part.
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Standard errors, clustered at school level, in parentheses.

the reform effect on perceived stress is significantly different from zero for both groups. Students with higher mathematical competence do not show significantly more internalizing mental health problems after the G8 reform, while students with lower Math competence do. The heterogeneity in mental health problems supports the finding that the reforms might burden some students more than other as the increase in grade repetitions Huebener and Marcus (2017) and the lacking benefit of additional instruction time for some ninth-graders after the reform (Huebener et al., in press) suggest. Well-being in school is only negatively affected for lower ability students. There seems to be no heterogeneity with respect to mathematical competence for general subjective well-being.

Books at home– Students from higher socioeconomic backgrounds might, in addition to usually performing better in school, also be better-supported by their parents and therefore be less affected by the G8 reform. By interacting the reform effect with a dummy indicating whether the student reports to have at least 200 books at home (comparison group reports less than 200 books at home) effect heterogeneity with respect to socioeconomic background is displayed. There is no effect heterogeneity for stress and both measures of well-being. However, students from higher socioeconomic backgrounds show significantly more mental health problems after the reform, while the effect for students with fewer books at home is not statistically significantly different from zero. Two possible explanations might trigger this result, which is surprising at first sight:

First, students of higher socioeconomic background might be more pressured not only pass the *Abitur*, but to do so with good or excellent grades. Second, the decreased costs in terms of time spent in school instead of working might be perceived as more valuable for students from lower SES backgrounds.

Migration background– Typically students from a migration background perform less well at school and therefore they might suffer more from the G8 reform. For this heterogeneity analysis the G8 effect was interacted with the dummy for having a migration background. The heterogeneity analyses yield statistically insignificant interaction terms. Surprisingly, the direction of the interaction terms always points to better outcomes for students with a migration background. For internalizing mental health problems and well-being in school, the negative reform effects are solely driven by students without migration background.

To sum up, the results suggest some heterogeneity in the reform effects. Interestingly students who would typically be seen as stemming from disadvantaged groups seem to suffer less from the reforms than their more privileged peers.

4.6 Robustness and extensions

In order to test whether the results attained above are sensible to the methodology used in the analyses, alternative specifications were analyzed. This section provides an overview of the results.[20]

Control for relative age across both cohorts – An alternative interpretation of the reform effect would be to consider it without the age difference between G8 and G9 students as part of the reform effect. If the increase in stress and mental health problems was mainly driven by age differences between the students, the reform effect should be substantially smaller, when age is centered over the entire sample instead of centered separately by cohort.[21] The resulting reform effects on stress and well-being in school increase rather than decrease when this adjustment is made, while the impact on internalizing mental health problems and general subjective well-being are only slightly affected (table 4.6, Panel A). Accordingly, the age difference does not seem to be the main driving force of the reform effect.

Omit sports participation from controls – Sports participation might also serve as outcome of the reforms since an increased amount of hours spent in school might lead to less time spent on leisure activities like doing sports regularly. My initial results are robust against the exclusion of sports participation from the set of control variables (table 4.6, Panel A).

Regression without controls – Controls are not needed for identification of the effect of G8 on stress, mental health and well-being since I assume assignment to the treatment random conditional on the policy reform. When the analyses are conducted without any additional controls, results for stress and both measures of well-being change only marginally and do not change their significance levels. The effects on internalizing mental health symptoms are reduced marginally thereby also reducing their statistical significance (table 4.6, Panel A). Nevertheless, these changes remain very small and therefore I do not consider these results as threat to the initial assumptions.

Using weights – Selectivity in the survey participation might bias the results if students did not agree to participate in the study in nonrandom ways. The design of the study does not allow to control for this potential bias directly. Nevertheless, the participating schools provided grade point lists for all students in the respective cohorts regardless of whether they participated in the study or not. An analysis of participation dependent on student and average school grades, as

[20] Tables A4.9 & A4.10 provide the results separately by gender.

[21] There is some evidence, that psychological problems increase with age for younger adolescents between grade 5 and 9 (Bilz, 2008, p. 111ff).

Table 4.6: Robustness checks

	Perceived stress		Mental health problems		Well-being school		Well-being life in general	
Panel A: Robustness checks using analysis sample (N=2306)								
Baseline model for reference								
Reform (G8)	0.314***	(0.049)	0.111**	(0.048)	−0.182	(0.112)	0.065	(0.070)
Control for relative age over both cohorts								
Reform (G8)	0.526***	(0.076)	0.152**	(0.073)	−0.467***	(0.173)	0.099	(0.121)
Omit sports participation from controls								
Reform (G8)	0.317***	(0.049)	0.113**	(0.048)	−0.187	(0.112)	0.058	(0.070)
No controls								
Reform (G8)	0.298***	(0.052)	0.088*	(0.047)	−0.128	(0.114)	0.080	(0.069)
Using weights								
Reform (G8)	0.272***	(0.053)	0.088*	(0.052)	−0.071	(0.114)	0.142*	(0.079)
Panel B: Robustness checks with additional sample restrictions								
Exclude grade repeaters (N=2192)								
Reform (G8)	0.316***	(0.051)	0.114**	(0.049)	−0.189	(0.114)	0.054	(0.071)
Exclude old G8 (N=2282)								
Reform (G8)	0.311***	(0.049)	0.106**	(0.048)	−0.187	(0.113)	0.074	(0.069)
Panel C: Using wave 2010/11 as control and wave 2012/13 as treatment (N=2406)								
Reform (G8)	0.438***	(0.048)	0.286***	(0.045)	−0.293***	(0.098)	−0.183***	(0.064)

Notes: NEPS BW D_3-0-0, wave 2011/12 (panel A and B) and waves 2010/11 and 2012/13 (Panel C). OLS regressions. The dependent variables are standardized to a mean of zero and a standard deviation of one. All models contain a constant and control additionally for age deviation from cohort median, previous grade repetition, sports participation, migration background, books at home and gender (only pooled models).
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Standard errors, clustered at school level, in parentheses.

well as school random effects shows some differences in participation dependent on (unobserved) school characteristics and for the G9 students also differences in participation dependent on the school wide average grade (Schönberger and Aßmann, 2014). To exclude that my results are driven by these selectivity issues, an additional specification using the weights provided by NEPS[22] shows slightly lower reform effects on all outcomes except satisfaction with life in general for a representative sample of students in Baden-Wurttemberg (table 4.6, panel A). The reform effect on satisfaction with life in general becomes significantly positive in this specification.

G8 as fast track program – To account for students who were in fast track programs for highly able students before the reform, the following adjustment was made: For each of the students in the last G9 cohort who graduated already in 2011 as part of the old fast track program, and therefore was sampled in the first wave of the Additional Study Baden-Wurttemberg, a student from the first G8 cohort (sampled in the second wave in 2012) was selected via nearest neighbor matching.[23] In table 4.6 (panel B) the summary of the resulting estimations shows that an exclusion of these matched students from the sample results only in marginal differences in effect sizes that do not change the signs or significance levels reported in the baseline models (panel

[22] More information on the weight generation can be found in Schönberger and Aßmann (2014).

[23] Students were matched on the following variables: migration background, age in months, female, sports participation, previous grade repetition, books at home, and mathematical competence. The student from the second wave with the smallest distance to the respective fast track student was selected. In case of ties, i.e. more then one student was at the lowest distance, all tied students were selected.

A, first line). This result is not particularly surprising, since the procedure led to the exclusion of only 26 students.

Exclusion of grade repeaters – Since, by design of the reform there are only very few grade repeaters in the G8 cohort and students who repeated a grade are likely to be different from other students in several other domains as well, I also conducted all analyses excluding students who repeated at least one grade during their schooling career. Excluding these students does not alter the results (cf. table 4.6, panel B).

Using waves 2010/11 and 2012/13 – Further, to see whether there are hints that the effects found here are only valid for the double cohort, I also estimated the reform effects for the cohorts immediately before and after the double cohort. In this case, the control group is the last pure G9 cohort which graduated in 2011 and the treatment group is the first pure G8 cohort which graduated in 2013. The results presented in table 4.6 (panel C) point towards larger reform effects in this setting. Most notably in this setting, males also have a significantly higher prevalence of internalizing mental health symptoms (table A4.9 in the appendix). Nevertheless, these results must be treated cautiously, as students in the treatment and control group faced different curricula even in the last two school years and also general time trends, if they exist, potentially have more impact on the results as the G8 cohort in this setting was born three years instead of one year after the G9 cohort they are compared to.

Different construction of dependent variables – I based the variable construction of the indexes for stress on the description of the index for internalizing mental health problems in Bergmüller (2007), because no documentation exists for the index of perceived stress. To show that the results are robust to different operationalizations of the dependent variables, I also estimated the reform effects allowing different amounts of missing values in the indexes, using separate standardization by gender, and using factor loads as weights for the items. None of these adjustments change the resulting reform effects more than marginally. Additionally, I also generated the dependent variables by counting how many items were answered with the highest or the highest and second highest answer categories and then standardizing this number to make results comparable. With this adjustment, the reform effects are slightly smaller. However, only in the model counting only the highest category the reform effect on stress for males turns insignificant.

4.7 Discussion

The current paper used data from the National Educational Panel Study (NEPS) to assess consequences on several measures of well-being induced by a recent German school policy reform that reduced school duration by one year while keeping cumulative instruction time constant. The results show that perceived stress and indicators of internalizing mental health problems increase for students in the first reformed cohort, while well-being in school is only slightly affected and general subjective well-being is not affected at all. There are substantive gender differences in the effects. While stress increases by 40 % of a standard deviation for females, the effect for males is substantially lower (16% of a standard deviation). Internalizing mental health problems increase and well-being in school decreases for females yet not for males.

In the short run, the G8 reform seems to have adverse effects on student mental health and perceived stress. Yet, the NEPS-data do not allow for a long-run evaluation. It is not possible to tell whether the effects persist over time. Observing the double cohort also means that the effects found might be at least in part due to implementation problems of the reform and might level off over time. Yet, the robustness analyses do not point into this direction. To assess the long-run consequences the analysis of another dataset would be needed.

One further caveat of analyzing the double cohort is that the first cohort of G8-students might suffer from the reform to a larger extent because they are directly compared to their G9 peers and

Table 4.7: Different calculations for dependent variables

	Stress			Mental health problems		
	Pooled	Male	Female	Pooled	Male	Female
Separate standardization by gender[a]						
Reform (G8)	0.319***	0.226***	0.394***	0.110**	0.011	0.183***
	(0.050)	(0.069)	(0.063)	(0.050)	(0.070)	(0.061)
Observations	2306	986	1320	2306	986	1320
Allow for no missing values[b]						
Reform (G8)	0.314***	0.198***	0.402***	0.122**	−0.006	0.213***
	(0.053)	(0.068)	(0.066)	(0.050)	(0.064)	(0.065)
Observations	2154	918	1236	2154	918	1236
No restriction for missings[c]						
Reform (G8)	0.316***	0.217***	0.393***	0.111**	0.013	0.182***
	(0.049)	(0.065)	(0.062)	(0.047)	(0.061)	(0.062)
Observations	2316	988	1328	2316	988	1328
Count all occurrences of largest and second largest value[d]						
Reform (G8)	0.302***	0.199***	0.383***	0.115**	0.015	0.188***
	(0.050)	(0.067)	(0.064)	(0.046)	(0.059)	(0.061)
Observations	2322	992	1330	2322	992	1330
Count all occurrences of largest value[e]						
Reform (G8)	0.258***	0.096	0.377***	0.094*	−0.029	0.184***
	(0.053)	(0.066)	(0.062)	(0.047)	(0.060)	(0.062)
Observations	2322	992	1330	2322	992	1330
Using factor loads[f]						
Reform (G8)	0.318***	0.216***	0.397***	0.097**	−0.049	0.205***
	(0.052)	(0.069)	(0.064)	(0.047)	(0.058)	(0.066)
Observations	2240	958	1282	2227	947	1280

Notes: NEPS BW D_3-0-0 wave 2011/12. OLS regressions. The dependent variables are standardized to a mean of zero and a standard deviation of one. All models contain a constant and control additionally for age deviation from cohort median, previous grade repetition, sports participation, migration background, books at home, gender (only pooled models).

[a] Since there are severe differences between male and female students in the outcomes, in addition to the normal standardization over the entire sample, the outcomes were standardized separately for males and females.

[b] In this specification, indexes for stress and mental health are only calculated if there is no missing value in any of the respective items.

[c] In this specification, indexes for stress and mental health are calculated as long as there is at least one nonmissing value in among the respective items.

[d] In this specification all occurrences of the highest and second highest answer category are counted. The number of occurrences is then standardized, in order to allow comparison of the results to the other specifications

[e] This specification is similar to the previous specification, except that it only counts occurrences of the highest possible answer category

[f] In this specification, before the standardization takes place, the items in each score are weighted by their predicted rotated factor loadings, which result from factor analysis.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Standard errors, clustered at school level, in parentheses.

have to compete with the G9 peers for limited resources after graduation (e. g. apprenticeships, volunteering activities, university places, student accommodation in university cities).

On the other hand do comparisons of the G9 students in wave 2010/11 who were in the last pure G9 cohort and the G9 students in double cohort rather point towards higher effects of the reform, as the stress and internalizing mental health problem scores are lower in the last pure G9 cohort. At the same time there are no significant differences in all control variables between the two waves (cf. table A4.4 in the appendix).

Whether or not general time trends impair the results is hard to tell. A placebo estimation using G9-students from wave 2010/11 and 2011/12 is not helpful, because also the G9-students in wave 2011/12 might be affected by being part of the double cohort. A placebo estimation using the first two G8-cohorts does, however not yield significant effects of the placebo-reform.[24] So at least there does not seem to be a increase in mental health problems and perceived stress in every year.

The reform effects in this analysis setup could also be driven by secular age trends in mental health problems if they were decreasing over age. In this case, more mental health problems would be expected in the G8-cohort which graduated at a younger age than the G9-cohort. Contrarily, most literature suggests that the prevalence of mental health problems (e. g. depression and anxiety, among others) increases with age at the onset of adolescence (Rushton et al., 2002; Garrison et al., 1989). The increase levels off either during late adolescence (Lewinsohn et al., 1998; Thapar et al., 2012) or during early adulthood (Patel et al., 2007; Merikangas et al., 2009, 2010). However, estimates of the levels at which disorders occur vary substantially and depend on the specific disorders taken into account as well as on the analyzed samples (Costello et al., 2011). Most students in my sample are aged 17–19. In this age group, students are more likely to face increasing or stable instead of decreasing frequencies of mental health problems. Therefore, the reform effects are unlikely to be driven by secular age-specific trends.

In how far results from Baden-Wurttemberg can be generalized to other states, is a matter for further discussion. In general, the education systems across the German Länder are comparable. Nevertheless, there are a few differences, for example does Baden-Wurttemberg have a comparatively low share of comprehensive schools, where students of different abilities are instructed together. This makes selection out of the reform harder for students who want to attain the *Abitur*.[25] The increase of weekly instruction hours in Baden-Wurttemberg was higher than in other states (Huebener et al., in press), this could mean that students are more affected by the reform, on the other hand it also means they had more instruction time than students from other states to master the curriculum, it is hard to tell without additional data what effect this difference might have. Another difference to other states is that Baden-Wurttemberg, among others, has a long tradition of central exit examinations, whereas central exit examinations were quite new in other states when they implemented the G8-reform. Since, most states introduced central exit examinations before the introduction of G8, most students in the first graduating G8 cohorts, regardless of state, had to face central exit examinations to attain their *Abitur*. Further, the double cohort in Baden-Wurttemberg was instructed together, which could potentially lead to spill-over effects between the cohorts. However, I do not expect results from Baden-Wurttemberg to be fundamentally different from other states, but since the dataset used in this study is restricted to Baden-Wurttemberg, the possibility of generalization to other German Länder cannot be tested.

The gender gap in internalizing mental health problems might be related to the finding from (Morin, 2015) who finds that the reform in Canada improved college grades and timely graduation for males relative to females and relates it to a larger literature showing that males respond

[24] Results are not reported but available upon request.

[25] Most comprehensive schools maintained G9 systems. Since students of all abilities are instructed together in the lower grade levels, this was necessary to meet the requirements for the *Abitur*.

differently to competition than females (e. g. Gneezy et al., 2003; Gneezy and Rustichini, 2004). In the observed reform in Baden-Wurttemberg, there has been at least some increase in competition in the double cohort. Yet, the results on the first G8 cohort does not imply a similar increase in competition but also shows larger effects for females. Another possible explanation also be that, for boys, preparation time for school decreases with increasing age, while it tends to increase for girls (Böhm-Kasper and Weishaupt, 2002, p. 481), therefore an increase in learning intensity would affect females more than males.

Another possible explanation might be that, according to Huebener and Marcus (2017), the reform increased male grade repetition rates in the final two years of school much more than female grade repetition rates. My results suggest that prior grade repetition is harmful to mental health. Furthermore, increased male grade repetition in the penultimate year of schooling could mean that the sample of G8 students in this analysis is positively selected. However, in this case I would most likely underestimate the reform effects. The stronger results in the specifications using the first pure G8 and last pure G9 cohorts might be interpreted as an indicator of such a selection effect.

While I consider an increase in learning intensity as the main determinant of the reform effect, the effect might also be influenced by changes in the curricula between G8 and G9, age differences, competition, and potential spill-over effects. As described in chapter 4.2.2, the curriculum of the double cohort was designed to minimize the impact of different curricula from grades 5 to 10/11. A minor role of different curricula can, however, not be ruled out in this analysis framework. In chapter 4.6, I show that age is not the only determinant of the reform effect: accounting for relative age across both cohorts rather increases the impact of the G8 reform on stress, while it does not change the effect on internalizing mental health problems. Finally spillover effects between the two cohorts cannot be ruled out as I only observe students who were instructed together. Future research might be able to further disentangle the composition of the reform effect.

In total, the effects of the reform show increases in perceived stress and internalizing mental health problem symptoms, which are mainly driven by females. Whether the results from this study can be extended to other cohorts and states remains to be shown in further studies. However, the results show that reforms like the one examined in this study can affect perceived stress and symptoms of internalizing mental health problems. In the light of recent developments in Germany to change school systems again, policymakers should keep in mind that this might impair students' mental health.

4.8 Appendix

Table A4.1: Description of variables

Variable	Description
Stress	Standardized index of perceived stress constructed from 15 items, mean 0 and standard deviation 1 (for contained items see table A4.2)
Mental health problems	Standardized index of internalizing mental health problems constructed from 15 items, mean 0 and standard deviation 1 (for contained items see table A4.3)
Well-being life	Self-reported satisfaction with life in general on a scale from 0 to 10
Well-being school	Self-reported satisfaction with school on a scale from 0 to 10
Reform (G8)	Dummy indicating whether student was affected by the reform
Female	Dummy indicating whether student is female
Age (years)	Student age in years
Age - median	Student age in months minus median age of respondents in own cohort
Repeated	Dummy indicating whether student ever repeated a grade
Migration background	Dummy indicating whether student has a migration background, i. e. at least one parent was not born in Germany
Books at home	Number of Books at home, categories: 0-100, 101-200, 201-500, more than 500 books
Sports	Dummy indicating whether student does sports on more than one day per week

Table A4.2: Perceived stress

How much do you agree with the following statements?	
1	I am tense when I get home from school.
2	Sometimes I have trouble falling asleep because problems from school are on my mind.
3	It happens that I react irritably when others start talking to me about school.
4	I feel that school is overwhelming me.
5	Even during my free time I think about troubles at school.
6	I consider the requirements at school in general as stressful.
7	After school I am often exhausted.
8	Thinking of school makes me feel uncomfortable.
9	Pressure at school is too high.
10	School is eating me up.
11	It is hard for me to conciliate school with other obligations.
12	School often makes me feel tired and exhausted.
13	It is easy for me to recover from school during my free time. (reversed)
14	I can relax well during my free time. (reversed)
15	Apart from school, I do not have time for anything else.
Response format: 1 = not at all true; 2 = hardly true; 3 = moderately true; 4 = exactly true	

Notes: NEPS BW D_3-0-0 wave 2011/12. Translation by NEPS.

Table A4.3: Symptoms of internalizing mental health problems

How often have you had the following problems in the last 6 weeks?	
1	Nervousness, inner anxiety
2	Headaches
3	Strong heart palpitations
4	Fear that it's all getting too much
5	Difficulty concentrating
6	Sleep disturbances
7	Bad dreams
8	Excessive sweating
9	Vomiting
10	Easily irritable
11	Feelings of dizziness
12	Tiredness, fatigue
13	Incapable of relaxing
14	Severe forgetfulness, distraction
15	Angry at everything
16	Feeling of being worthless
17	Fear of going to school
18	Shakiness, weakness
19	Nausea
20	Loss of appetite
21	Backache
22	Sadness
23	Feeling that excessive demands are being made of me
24	Eating binges
25	Feeling of inner emptiness
26	Stomach ache
Response format: 1 = never; 2 = 1-2 times; 3 = 3-6 times; 4 = more than 6 times	

Notes: NEPS BW D_3-0-0 wave 2011/12. Translation by NEPS.

Table A4.4: Differences in variables between G9-students in wave 2010/11 and G9-students in wave 2011/12

	Mean		**Equality of means**	
	Wave 2010/11	Wave 2011/12	Difference	t-stat
Stress score	1.929	2.015	−0.0859	−3.851***
Mental health problem score	1.800	1.869	−0.0693	−3.227**
Well-being school	7.251	7.129	0.122	1.470
Well-being life	7.763	7.701	0.0615	0.909
Female	0.556	0.568	−0.0122	−0.596
Age (years)	18.47	18.40	0.0769	3.272**
Repeated a grade	0.104	0.0853	0.0190	1.573
Sports	0.790	0.778	0.0121	0.711
Migration backgr.	0.235	0.222	0.0138	0.800
0-100 books at home %	0.189	0.179	0.0102	0.636
101-200 books at home %	0.163	0.162	0.000702	0.0462
201-500 books at home %	0.316	0.311	0.00515	0.270
500 + books at home %	0.328	0.346	−0.0176	−0.906
Observations	1198	1160		

Notes: NEPS BW D_3-0-0 wave 2010/11 and 2011/12

Table A4.5: Differences in variables between G8-students in wave 2011/12 and G8-students in wave 2012/13

	Mean		Equality of means	
	Wave 2011/12	Wave 2012/13	Difference	t-stat
Stress score	2.193	2.201	−0.008	−0.317
Mental health problem score	1.920	1.972	−0.052	−2.259*
Well-being school	6.998	6.926	0.0725	0.861
Well-being life	7.775	7.571	0.204	3.042**
Female	0.577	0.568	0.00926	0.453
Age (years)	17.33	17.45	−0.118	−5.499***
Repeated a grade	0.015	0.110	−0.095	−9.732***
Sports	0.763	0.736	0.0262	1.465
Migration backgr.	0.209	0.240	−0.031	−1.775
0-100 books at home %	0.191	0.179	0.012	0.740
101-200 books at home %	0.172	0.151	0.021	1.393
201-500 books at home %	0.307	0.311	−0.004	−0.199
500 + books at home %	0.328	0.358	−0.030	−1.534
Observations	1206	1208		

Notes: NEPS BW D_3-0-0 wave 2011/12 and 2012/13

Table A4.6: Differences in variables between G9-students in wave 2010/11 and G8-students in wave 2012/13

	Mean		Equality of means	
	Wave 2010/11	Wave 2012/13	Difference	t-stat
Stress score	1.929	2.201	−0.272	−11.84***
Mental health problem score	1.800	1.972	−0.172	−7.875***
Well-being school	7.251	6.926	0.325	3.935***
Well-being life	7.763	7.571	0.192	2.908**
Female	0.556	0.568	−0.0121	−0.594
Age (years)	18.47	17.45	1.021	39.30***
Repeated a grade	0.104	0.110	−0.00535	−0.422
Sports	0.790	0.736	0.0533	3.057**
Migration backgr.	0.235	0.240	−0.00440	−0.252
0-100 books at home %	0.189	0.179	0.0101	0.632
101-200 books at home %	0.163	0.151	0.0123	0.821
201-500 books at home %	0.316	0.311	0.00514	0.270
500 + books at home %	0.328	0.358	−0.0299	−1.537
Observations	1198	1176		

Notes: NEPS BW D_3-0-0 wave 2010/11 and 2012/13

Table A4.7: Effect heterogeneity: Stress and mental health problems

	Stress			Mental health problems		
	Pooled	Male	Female	Pooled	Male	Female
Baseline model for reference						
Reform (G8)	0.314***	0.214***	0.392***	0.111**	0.010	0.186***
	(0.049)	(0.065)	(0.063)	(0.048)	(0.061)	(0.063)
Age higher than median						
Reform (G8)	0.278***	0.165**	0.359***	0.056	−0.036	0.121
	(0.054)	(0.074)	(0.075)	(0.055)	(0.073)	(0.076)
Interaction	0.075	0.096	0.072	0.113	0.090	0.141
	(0.062)	(0.095)	(0.093)	(0.068)	(0.101)	(0.093)
Wald test p-value[a]	0.000***	0.003***	0.000***	0.009***	0.523	0.002***
Math competence above median[b]						
Reform (G8)	0.373***	0.245**	0.441***	0.151**	−0.054	0.248***
	(0.050)	(0.101)	(0.055)	(0.067)	(0.110)	(0.071)
Interaction	−0.118	−0.027	−0.139	−0.075	0.114	−0.160
	(0.075)	(0.117)	(0.111)	(0.085)	(0.126)	(0.117)
Wald test p-value[a]	0.001***	0.006***	0.009***	0.225	0.381	0.397
More than 200 books at home (higher SES)						
Reform (G8)	0.265***	0.169*	0.352***	0.072	−0.067	0.180*
	(0.064)	(0.092)	(0.083)	(0.079)	(0.089)	(0.105)
Interaction	0.075	0.070	0.061	0.061	0.121	0.008
	(0.076)	(0.118)	(0.099)	(0.091)	(0.118)	(0.116)
Wald test p-value[a]	0.000***	0.006***	0.000***	0.022**	0.507	0.009***
Students with migration background						
Reform (G8)	0.341***	0.251***	0.415***	0.137***	0.048	0.208***
	(0.057)	(0.079)	(0.070)	(0.048)	(0.063)	(0.063)
Interaction	−0.127	−0.185	−0.103	−0.124	−0.188	−0.100
	(0.102)	(0.162)	(0.143)	(0.095)	(0.141)	(0.135)
Wald test p-value[a]	0.016**	0.607	0.019**	0.890	0.297	0.424
Observations	2306	986	1320	2306	986	1320

Notes: NEPS BW D_3-0-0 wave 2011/12. OLS regressions. The dependent variables are standardized to a mean of zero and a standard deviation of one. All models contain a constant and control additionally for age deviation from cohort median, previous grade repetition, sports participation, migration background, books at home, gender (only pooled models), and a dummy indicating whether the student's mathematical competence is above the median in her cohort (only mathematical competence part). The interaction term displays the interaction of the reform with being older than the median student, with above median mathematical competence, with having more than 200 books at home or with having a migration background. [a]Test of the hypothesis H_0 Reform (G8) + Interaction = 0. [b]Due to missing data, 3 observations are missing in the mathematical competence part.
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Standard errors, clustered at school level, in parentheses.

Table A4.8: Effect heterogeneity: Well-being in school and life

	School			Life in General		
	Pooled	Male	Female	Pooled	Male	Female
Baseline model for reference						
Reform (G8)	−0.182	−0.082	−0.243*	0.065	0.143	0.002
	(0.112)	(0.155)	(0.130)	(0.070)	(0.095)	(0.093)
Age higher than median						
Reform (G8)	−0.226*	−0.174	−0.261	0.028	0.151	−0.059
	(0.133)	(0.174)	(0.157)	(0.088)	(0.105)	(0.109)
Interaction	0.091	0.181	0.040	0.077	−0.015	0.133
	(0.135)	(0.177)	(0.189)	(0.118)	(0.158)	(0.162)
Wald test p-value[a]	0.299	0.973	0.187	0.273	0.335	0.595
Math competence above median[b]						
Reform (G8)	−0.236*	−0.106	−0.283*	0.015	0.124	−0.051
	(0.121)	(0.255)	(0.146)	(0.108)	(0.168)	(0.130)
Interaction	0.109	−0.007	0.145	0.099	0.027	0.134
	(0.170)	(0.314)	(0.209)	(0.146)	(0.189)	(0.202)
Wald test p-value[a]	0.423	0.550	0.462	0.229	0.159	0.566
More than 200 books at home (higher SES)						
Reform (G8)	−0.185	−0.105	−0.215	0.066	−0.046	0.140
	(0.202)	(0.295)	(0.229)	(0.126)	(0.169)	(0.163)
Interaction	0.005	0.036	−0.041	−0.001	0.298	−0.209
	(0.209)	(0.336)	(0.233)	(0.152)	(0.239)	(0.174)
Wald test p-value[a]	0.115	0.687	0.055*	0.446	0.068*	0.483
Students with migration background						
Reform (G8)	−0.237*	−0.064	−0.360**	0.055	0.153	−0.027
	(0.119)	(0.164)	(0.143)	(0.070)	(0.104)	(0.089)
Interaction	0.259	−0.088	0.537**	0.048	−0.050	0.132
	(0.248)	(0.453)	(0.232)	(0.196)	(0.274)	(0.236)
Wald test p-value[a]	0.927	0.719	0.423	0.587	0.681	0.653
Observations	2306	986	1320	2306	986	1320

Notes: NEPS BW D_3-0-0 wave 2011/12. OLS regressions. The dependent variables are on an eleven-point Lickert-type scale. All models contain a constant and control additionally for age deviation from cohort median, previous grade repetition, sports participation, migration background, books at home, gender (only pooled models), and a dummy indicating whether the student's mathematical competence is above the median in her cohort (only mathematical competence part). The interaction term displays the interaction of the reform with being older than the median student, with above median mathematical competence, with having more than 200 books at home or with having a migration background. [a]Test of the hypothesis H_0 Reform (G8) + Interaction = 0. [b]Due to missing data, 3 observations are missing in the mathematical competence part.
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Standard errors, clustered at school level, in parentheses.

Table A4.9: Robustness checks: Stress and mental health problems

	Stress			Mental health problems		
	Pooled	Male	Female	Pooled	Male	Female
Panel A: Robustness checks using analysis sample						
Baseline model for reference						
Reform (G8)	0.314***	0.214***	0.392***	0.111**	0.010	0.186***
	(0.049)	(0.065)	(0.063)	(0.048)	(0.061)	(0.063)
Control for relative age over both cohorts						
Reform (G8)	0.526***	0.573***	0.474***	0.152**	0.084	0.190*
	(0.076)	(0.115)	(0.093)	(0.073)	(0.098)	(0.095)
Omit sports participation from controls						
Reform (G8)	0.317***	0.218***	0.395***	0.113**	0.013	0.187***
	(0.049)	(0.064)	(0.065)	(0.048)	(0.060)	(0.063)
No controls						
Reform (G8)	0.298***	0.198***	0.368***	0.088*	0.002	0.146**
	(0.052)	(0.061)	(0.065)	(0.047)	(0.056)	(0.061)
Separate standardization by gender						
Reform (G8)	0.319***	0.226***	0.394***	0.110**	0.011	0.183***
	(0.050)	(0.069)	(0.063)	(0.050)	(0.070)	(0.061)
Using weights						
Reform (G8)	0.272***	0.190**	0.339***	0.088*	−0.003	0.154**
	(0.053)	(0.074)	(0.068)	(0.052)	(0.066)	(0.072)
Observations	2306	986	1320	2306	986	1320
Panel B: Robustness checks with additional sample restrictions						
Exclude grade repeaters						
Reform (G8)	0.316***	0.225***	0.388***	0.114**	0.006	0.198***
	(0.051)	(0.068)	(0.065)	(0.049)	(0.063)	(0.062)
Observations	2192	932	1260	2192	932	1260
Exclude old G8						
Reform (G8)	0.311***	0.207***	0.393***	0.106**	0.002	0.183***
	(0.049)	(0.063)	(0.062)	(0.048)	(0.058)	(0.063)
Observations	2282	978	1304	2282	978	1304
Panel C: Using wave 2010/11 as control and wave 2012/13 as treatment						
Reform (G8)	0.436***	0.403***	0.464***	0.282***	0.254***	0.306***
	(0.049)	(0.050)	(0.067)	(0.045)	(0.043)	(0.064)
Observations	2419	1055	1364	2419	1055	1364

Notes: NEPS BW D_3-0-0, wave 2011/12 (panel A and B) and waves 2010/11 and 2012/13 (Panel C). OLS regressions. The dependent variables are standardized to a mean of zero and a standard deviation of one. All models contain a constant and control additionally for age deviation from cohort median, previous grade repetition, sports participation, migration background, books at home and gender (only pooled models).

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Standard errors, clustered at school level, in parentheses.

Table A4.10: Robustness checks: Well-being in school and life

	Well-being School			Well-being Life		
	Pooled	Male	Female	Pooled	Male	Female
Panel A: Robustness checks using analysis sample						
Baseline model for reference						
Reform (G8)	−0.182	−0.082	−0.243*	0.065	0.143	0.002
	(0.112)	(0.155)	(0.130)	(0.070)	(0.095)	(0.093)
Control for relative age over both cohorts						
Reform (G8)	−0.467***	−0.346	−0.553**	0.099	0.060	0.134
	(0.173)	(0.246)	(0.226)	(0.121)	(0.198)	(0.162)
Omit sports participation from controls						
Reform (G8)	−0.187	−0.091	−0.245*	0.058	0.131	−0.001
	(0.112)	(0.153)	(0.131)	(0.070)	(0.094)	(0.095)
No controls						
Reform (G8)	−0.128	−0.044	−0.189	0.080	0.139	0.037
	(0.114)	(0.151)	(0.138)	(0.069)	(0.085)	(0.095)
Using weights						
Reform (G8)	−0.071	0.040	−0.131	0.142*	0.213**	0.085
	(0.114)	(0.155)	(0.134)	(0.079)	(0.104)	(0.101)
Observations	2306	986	1320	2306	986	1320
Panel B: Robustness checks with additional sample restrictions						
Exclude grade repeaters						
Reform (G8)	−0.189	−0.071	−0.267**	0.054	0.137	−0.015
	(0.114)	(0.160)	(0.130)	(0.071)	(0.099)	(0.093)
Observations	2192	932	1260	2192	932	1260
Exclude old G8						
Reform (G8)	−0.187	−0.091	−0.246*	0.074	0.165*	0.003
	(0.113)	(0.156)	(0.131)	(0.069)	(0.094)	(0.091)
Observations	2282	979	1303	2282	979	1303
Panel C: Using wave 2010/11 as control and wave 2012/13 as treatment						
Reform (G8)	−0.293***	−0.250	−0.317***	−0.183***	−0.170*	−0.190**
	(0.098)	(0.149)	(0.108)	(0.064)	(0.099)	(0.086)
Observations	2406	1047	1359	2406	1047	1359

Notes: NEPS BW D_3-0-0, wave 2011/12 (panel A and B) and waves 2010/11 and 2012/13 (Panel C). OLS regressions. All models contain a constant and control additionally for age deviation from cohort median, previous grade repetition, sports participation, migration background, books at home and gender (only pooled models). Dependent variables: 'How satisfied are you with your situation at school?' and 'How satisfied are you currently and in general terms, with your life?' (11-point scales, translated by NEPS).
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Standard errors, clustered at school level, in parentheses.

Chapter 5

Health Effects of Instruction Intensity Evidence from a Natural Experiment in German High-Schools

Johanna Sophie Quis and Simon Reif

5.1 Introduction

A large literature in health and education economics establishes a link between education and health. Although the correlation between education and different health outcomes is generally large, causal estimates of the relation are less conclusive (Montez and Friedman, 2015; Grossman, 2015).[1] From a theoretical perspective, Grossman (1972) suggested that education promotes abilities that increase health production and also marginal returns to health inputs.

A common strategy to establish a causal link between education and health is to use changes in compulsory schooling laws as an exogenous increase in years of education. In her seminal work, Lleras-Muney (2005) finds a reduction in mortality when education increases.[2] Since then, increases in compulsory schooling have been used to also analyze the effects of education on other health outcomes. Self-assessed health seems to increase (Oreopoulos, 2006; Arendt, 2005; Kemptner et al., 2011; Mazzonna, 2014) and obesity is reduced (Arendt, 2005; Kemptner et al., 2011; Brunello et al., 2013) when students stay in school longer. Crespo et al. (2014) find that more schooling increases mental health and Mazzonna (2014) finds that it reduces depression for males while Dursun and Cesur (2016) show that subjective well-being of women increased when they attained at least a middle-school degree. In contrast, Avendanoy et al. (2017) show in a recent study that increased compulsory schooling has negative effects on female mental health later in life.

Although these quasi experimental studies are appealing because years of education increase exogenously, the local average treatment effects do not necessarily capture the effect of more education in the Grossmann framework, which focuses on abilities acquired in school. The underlying assumption in most studies is that additional years of education increase abilities in the health production function. This assumption, however, does not necessarily need to apply. There are at least two opposing channels in which more years of schooling can influence health independent of ability acquirement: Negative effects of involuntary schooling and positive effects of health promotion in schools.

[1] Both overview articles are a good summary of studies on both correlations and causal effects between education and health.

[2] Most studies that also analyze mortality using mandatory school year increases generally find smaller but still positive effects (Van Kippersluis et al., 2011; Meghir et al., 2012; Fischer et al., 2013; Gathmann et al., 2015) however sometimes insignificant (Albouy and Lequien, 2009; Mazumder, 2008). One exception are Clark and Royer (2013) who find no effect.

Involuntary participation in schooling might have negative health effects, especially on mental health. Elias (1989) provides a review of early psychological studies on mental health of students and concludes that stress in school is a key factor detrimental to students mental health – a relation also found in more recent studies (Raufelder et al., 2013; Scrimin et al., 2016). Additionally, Forbes et al. (2017) show that less free time to produce health inputs generally decreases people's utilities. Stress in school can have long term negative effects as it can lead to reduced mental health later in life (Patton et al., 2014). It also increases the probability of being overweight in adolescence (Torres and Nowson, 2007) which has been shown to translate into various negative health effects later in life (Reilly and Kelly, 2011).

Schools do however also play an important role in promoting health. First, they can explicitly teach healthy behaviors. Such interventions have been successful in promoting physical activity but less in promoting healthy diets (Dzewaltowski et al., 2009; Sallis et al., 2003). Additionally, schools can provide facilities that support physical activities, and thereby promote students' health (Cradock et al., 2007). Strict enforcement policies and intervention programs in schools can also reduce smoking rates (Thomas et al., 2013). Further, Frisvold and Golberstein (2011) show that high school quality (measured i. e. by length of school year) does have a strong positive effect on students' health.

It is therefore an open question whether more years of schooling actually improve health. We contribute to this discussion by analyzing a natural experiment in Germany where 13 of the 16 federal states (*Bundesländer*) implemented a reform to shorten the academic track of secondary school (*Gymnasium*) from nine to eight years without changing the overall curricular content and overall instruction hours by increasing instruction intensity. The reform took place at different points in time affecting students who graduated between 2007 and 2016 depending on where they went to school. As only academic track students are affected, we can implement a triple difference-in-differences estimation strategy. We conduct our analysis with data from the German Socio-Economic Panel (SOEP).

The reform provides a well-suited setting to reinvestigate the relationship between more years of schooling and health. It did reduce years of schooling, but the curricular content remained unchanged. Therefore, the health effects of the reform are a priori ambiguous. Higher instruction intensity might provoke stress and reduce mental health. However, the same content is taught to students, so acquisition of abilities for healthy behavior should not have changed. Additionally, the reform increased students' flexibility because they left school one year earlier. To disentangle the direct effects of increased instruction intensity from the indirect effects of more flexibility we estimate the effects of the reform for a sample of students still in school and for a sample of recent graduates.

We employ different health measures to capture the effect on three health dimensions: The widely used subjective measure of self assessed health, body mass index (BMI) as a quasi objective measure and indicators for mental well-being. Self-assessed health is not affected by the reform, neither for students in school nor for graduates. We find that the reform increased BMI and reduced mental well-being for women in school. After graduation the effect is reversed: BMI is lower and mental well-being is higher for women after they finished school. Males are not affected by the reform. Our results suggest that while the reduction in years increases stress in school, it also increases flexibility for students earlier in life, facilitating life choices that improve health.

The rest of this chapter is structured as follows: In the next section we provide details on the school reform we exploit, our identification strategy and the data we use. We present our results in section 5.3 and discuss them in section 5.4.

5.2 Data and Empirical Strategy

5.2.1 The German Secondary Education Reform (G8)

In Germany, students are usually divided into three secondary schooling tracks after four years of elementary school.[3,4] Two vocational tracks (*Hauptschule* and *Realschule*) prepare students for vocational training, which starts after grade 9 or 10. In the academic track (*Gymnasium*), students are prepared to go to university. Some states also have comprehensive schools (*Gesamtschule*) where students are not split between tracks. Between 2003 and 2007, 13 of the 16 German states reduced the duration of the academic track from nine to eight years, resulting in a decrease of total school years from 13 to 12. The main motivation for this reform was to reduce students' age when they enter university and the labor market to a level comparable with other European countries. While high school duration was reduced by one year, the course content and total hours of instruction had to remain constant to satisfy federal regulations. Instruction hours exceeding the requirements were mostly abolished during the reform, reducing total instruction hours by 2,6%. This reduction of excess hours means that average weekly instruction time increased by only 9.6% in contrast to the 12.5% increase that would have resulted from fitting nine into eight years.[5]

Table 5.1: Schedule of Reform Implementation in German States

State	**Implementation**	**First 12 year grad.**
Saxony-Anhalt	2003	2007
Mecklenburg-Vorpommern	2004	2008
Saarland	2001	2009
Hamburg	2002	2010
Bavaria[a]	2004	2011
Lower Saxony[a]	2004	2011
Baden-Württemberg[b]	2004	2012
Bremen	2004	2012
Berlin	2006	2012
Brandenburg	2006	2012
North Rhine-Westphalia[b]	2005	2013
Schleswig-Holstein[b]	2007	2016

[a] Bavaria and Lower Saxony are currently reintroducing general 9-year systems with an option to finish after 8 years if students are performing particularly well. In Lower Saxony, the first 9-year cohorts will graduate in 2021 (goo.gl/FMofr5). In Bavaria, the first 9-year-cohort will start grade 5 in 2017 (goo.gl/7ltypS).

[b] Baden-Württemberg, North Rhine-Westphalia, and Schleswig-Holstein are moving away from a pure 8-year system towards a parallel 8- and 9-year system (goo.gl/fa9Izs).

Table 5.1 gives an overview of the time frame of the reform for each state we use in our analysis.[6] The reform constitutes a well-suited natural experiment in two ways. First, it was

[3] In a few states the separation either used to take place or still does take place after grade 6.

[4] States are the administrative level at which educational policies are determined. Nevertheless, there is a federal commission, the *Kultusministerkonferenz*, which determines the framework of the German education system.

[5] This number results from evenly distributing the instruction time from the final school year over the eight previous years, deducting the 2.6% of instruction time that were removed. The increase in weekly instruction hours was larger in grades 7-10 (+3.75 hours) than in grades 5-6 (+2 hours) and 11-12 (+2.5 hours). Source: Homuth (2017), p. 25, own calculations.

[6] We exclude four of the 16 German states: In Hesse students had a long transition period where they could select into either eight or nine year academic track schools which hampers our identification strategy. Rhineland-

implemented in different states at different points in time only for one type of secondary school, which allows us to use a triple difference-in-differences estimation strategy. That is, we can compare academic track students who experience 13 years of schooling to those with 12 years of schooling and then compare them to vocational track students, who were not affected by the reform. Second, the assignment to the reform group can be assumed to be random as it would have been costly to avoid the reform–either by moving to another state or by choosing the vocational track with significantly lower expected lifetime earnings. Huebener and Marcus (2017) indeed show that the reform did not induce changes in the student population.

In recent years, several studies analyzed the effects of the reform, mainly from an education economics point of view. Büttner and Thomsen (2015) find that Math grades at graduation are worse for students who experienced increased instruction intensity, while grades in German are not affected. Huebener and Marcus (2017) find decreased grade point averages at graduation. Huebener et al. (in press) assess student competences at age 15 and find increased performance across all domains especially for highly skilled students. Dahmann (in press) examines cognitive skills at age 17 and at graduation and finds, in line with the two previously mentioned studies, higher numerical skills for males at age 17 and lower reasoning skills for both genders at graduation.[7] Hübner et al. (2017) find negative reform effects on English reading competence and no effects on mathematics and physics competence. Personality of students seems to be only marginally affected by the reform (Thiel et al., 2014; Dahmann and Anger, 2014). Students repeat grades more frequently (Huebener and Marcus, 2017) and delay (Meyer and Thomsen, 2016) or abandon (Marcus and Zambre, forthcoming) university enrollment. In cross sectional samples of first year university students, Kühn (2014) and Dörsam and Lauber (2015) do not find any difference in performance (which, however, could be driven by selection). 8-year students are more burdened by learning and report less time for jobs and volunteer work (Meyer and Thomsen, 2015)

There are four studies that evaluated the reform from a health related point of view. In an early cross sectional survey of students from one German city who were in grade 10 (8-year system) and 11 (9-year system) at the time of the interviews, Milde-Busch et al. (2010) do not find any health differences between reform and control group. Westermaier (2016) analyzes whether the reform led students to increase consumption of illegal drugs but does not find any effect. Most closely related to our study are Quis (fortcoming) and Hofmann and Mühlenweg (2017), who find weakly negative health effects of the reform. Quis (2015) compares the first 8-year graduating cohort to the last 9-year graduating cohort in Baden-Württemberg and finds an increase in perceived stress and symptoms of internalizing mental health problems for females, but no effect on subjective well-being.[8] Hofmann and Mühlenweg (2017) evaluate a pooled sample of students and graduates resulting in a slight decrease of mental health, but no effect on physical health or smoking behavior.

5.2.2 Estimation Strategy

We evaluate the reform effects for measures of three different health dimensions. First, the commonly used subjective measure of self assessed health, second BMI as an objective measure and third indicators for mental well-being which are whether students worry a lot and a standardized mental health measure for graduates. To identify the effect of increased instruction intensity on

---

Palatinate always had 8.5 years of academic track duration and is therefore a special case. Thuringia and Saxony always had eight years of academic track, so that there was no change in instruction intensity.

7 These results may seem paradox at first sight, but since 8-year students of a fixed age will have received more instruction time than 9-year students at the same age, they should perform better in order to catch up the missing year until graduation.

8 These results are also confirmed by Hübner et al. (2017).

students' health in a simple setting with treated (8y) and untreated (9y) regions, one would estimate the treatment effect using a standard difference-in-differences estimator from the average health levels:

$$ATE = (\bar{Y}^{8y}_{pre} - \bar{Y}^{8y}_{post}) - (\bar{Y}^{9y}_{pre} - \bar{Y}^{9y}_{post}), \tag{5.1}$$

which then can be estimated parametrically by

$$Y_i = ATE(8y_i * post_i) + \beta_1 8y_i + \beta_2 post_i + \boldsymbol{X\beta} + \epsilon_i, \tag{5.2}$$

where **X** is either a vector of ones or a matrix of additional covariates. In our setting, states switch from having 9 years to having 8 years, so instead of dummy variables for treated and untreated regions and time periods, the ATE can here be derived by controlling for a maximum set of state and time dummys ($\boldsymbol{S}$ and $\boldsymbol{T}$) and including a pseudo interaction-term *8years* which is one if a student went to school in a state *s* at time *t* when the 8 year regime was in place and zero otherwise. The coefficient δ is then the estimate for our ATE:

$$Y_i = \delta 8years_i + \boldsymbol{S\gamma} + \boldsymbol{T\kappa} + \boldsymbol{X\beta} + \epsilon_i. \tag{5.3}$$

Only students in the academic track of secondary school were affected by the reform. We can therefore use secondary school students in the non-academic tracks as a further control group in a triple difference-in-differences design. Interacting the treatment dummy *8years* as well as time and state fixed effects from equation (5.3) with an indicator for being in the academic track A_i leads to our main specification:

$$Y_i = \delta(8years_i \times A_i) + \alpha A_i + \boldsymbol{S}\gamma_1 + \boldsymbol{T}\boldsymbol{\kappa}_1 + (\boldsymbol{S} \times A_i)\boldsymbol{\gamma_2} + (\boldsymbol{T} \times A_i)\boldsymbol{\kappa_2} + \boldsymbol{X\beta} + \epsilon_i. \tag{5.4}$$

We use this model to estimate the effects first for students when they are 17. This means students are still in school but have several years of experience in school which gives us a direct reform effect. Our non academic-track control group consists of students in the vocational and comprehensive schools with the exception of the lower level *Hauptschule* because students usually leave this school at the age of 15.

In addition, we estimate equation (5.4) for graduates who finished school at least one year ago to obtain the indirect effect of the reform on health.

5.2.3 Health Indicators and Sample Selection

We conduct our analysis with data from the German Socio-Economic Panel (SOEP). It is a large, representative household panel in Germany that started in 1984 (Wagner et al., 2007; SOEP, 2015). To estimate the reform effect for students who are still in school, we use the youth study, sampling all children in SOEP households who turn 17 in the respective survey year. When we analyze reform effects after graduation, we use the personal interviews for the years 2008, 2010, 2012 and 2014, because they are the most recent waves that contain a broader set of health variables.

Health Indicators

Our first dependent variable is self assessed health on a scale from *very good* (1) to *very bad* (5). Although it is a subjective measure, it is widely used and has been shown to serve as a good proxy for more objective health measures (Idler and Benyamini, 1997; Van Doorslaer and Gerdtham, 2003).[9]

[9] It is however a controversial health measure as self assessed health varies with income and salience of personal health even if underlying health remains unchanged (Etilé and Milcent, 2006; Crossley and Kennedy, 2002). We

BMI is our second dependent variable. As a ratio of bodyweight and -height, a high BMI is a reliable indicator for overweight which has been shown to lead to various health problems and increase the risk of all-cause mortality (Global BMI Mortality Collaboration, 2016). In the SOEP, BMI is constructed from self-reported bodyweight and -height and especially self-reported weight might suffer from reporting error. This reporting error would however only be a threat to our identification strategy if the reporting error were correlated with our reform, which we have no reason to assume.

Our third health dimension is mental well-being. Here we have to use two different measures for students and for graduates. The only proxy for mental well-being available at age 17 is one answer from the Big-5 inventory. Here, students are asked whether they consider themselves as persons who worry a lot on a scale from *not at all* (1) to *very much* (7). Although Dahmann and Anger (2014) and Thiel et al. (2014) analyze the Big-5 indicators and do not find that the reform affected personality, we regard this question as a good proxy for mental well-being because it can be seen as a state varying indicator (Schutte et al., 2003).

For our sample of graduates we have access to a more objective measure of mental health, the *mental component scale* (MCS). It is a standardized compound measure of mental well-being constructed from the "SF-12v2^{TM} Health Survey", normalized to mean 50 and standard deviation 10 (Andersen et al., 2007) and has been shown to be a valid measure of mental health (Gill et al., 2007).

Sample Selection

We identify students affected by the reform from their year of school entry and state of residence in the year they turn 17. We restrict our sample to those students who have never repeated a grade.[10] For those students who did not state their year of school entry we impute it from month of birth.[11] Additionally, we drop students who are extreme outliers in terms of their reported BMI.[12] After list-wise deletion of students with missing data, we observe a total of 1274 students in school of whom 685 visited the academic track and 403 experienced increased instruction intensity.

Only every other wave of the SOEP includes detailed health related questions. For our analysis of graduates we therefore use the earliest wave with health variables available, which, at the earliest, was conducted in the year after students finished school. Students who did not graduate from the academic track (and hence graduated from vocational tracks earlier) are assigned a hypothetical graduation year for if they had been in the academic track. Students who graduated[13] in 2007 or before are assessed on their responses from the 2008 wave. If they graduated in 2008 or 2009, they are included in our sample with their responses from 2010. The same logic is applied for students who graduated in 2010 (2012) and 2011 (2013) who we observe in 2012 (2014). After graduation we observe 1387 individuals of whom 461 graduated from academic track schools and 134 experienced increased instruction intensity.

nevertheless include it in our set of dependent variables because it is a widely used measure, especially in the literature on education and health.

10 This restriction is necessary to ensure that grade repetitions do not bias the allocation to treatment or control group.

11 Usually children enter school in year t if they are six years old by August of year t. For those students for whom we observe school entry year the imputation is correct for 85% of the respective students. When further narrowing down to cases where the allocation to treatment or control group might be harmed due to the imputation, the comparison between imputation and known year of school entry yields a wrong allocation of 8% of the students around the cutoff. Assuming that we are also wrong in 8% of the cases where we cannot observe the real year of school entry (102 students) we would expect a total of 8.2 individuals being allocated to the wrong group.

12 We excluded students whose BMI differed more than 2.5 interquartile distances from the gender specific samples first or third quartile.

13 Or, in the case of vocational track students would have graduated, had they visited the academic track.

5.2.4 Descriptive Statistics

Table 5.2: Youth Sample: Descriptive Statistics

	Pooled	9 Years	8 Years	9 Years - 8 Years	Academic	Vocational	Academic - Vocational
	Mean (S.D.)	Mean (S.D.)	Mean (S.D.)	Difference (P-value)	Mean (S.D.)	Mean (S.D.)	Difference (P-value)
Self-assessed health	1.63	1.67	1.55	0.12**	1.60	1.68	−0.08*
	(0.71)	(0.71)	(0.67)	(0.03)	(0.69)	(0.74)	(0.06)
Body mass index	21.27	20.99	21.10	−0.11	21.06	21.52	−0.47***
	(2.64)	(2.43)	(2.60)	(0.56)	(2.53)	(2.74)	(0.00)
Worry a lot	4.36	4.24	4.42	−0.17	4.35	4.37	−0.02
	(1.70)	(1.65)	(1.65)	(0.17)	(1.65)	(1.75)	(0.85)
Female	0.51	0.52	0.51	0.01	0.52	0.50	0.01
	(0.50)	(0.50)	(0.50)	(0.73)	(0.50)	(0.50)	(0.61)
Age (months)	201.41	201.60	201.46	0.14	201.52	201.28	0.24
	(4.03)	(3.82)	(4.10)	(0.64)	(3.99)	(4.08)	(0.30)
Non-intact family	0.24	0.19	0.21	−0.03	0.20	0.29	−0.09***
	(0.43)	(0.39)	(0.41)	(0.41)	(0.40)	(0.45)	(0.00)
Migration background	0.19	0.15	0.19	−0.04	0.17	0.22	−0.05**
	(0.40)	(0.35)	(0.39)	(0.13)	(0.38)	(0.42)	(0.03)
High parental education	0.35	0.54	0.48	0.06	0.51	0.16	0.34***
	(0.48)	(0.50)	(0.50)	(0.14)	(0.50)	(0.37)	(0.00)
Rural	0.33	0.25	0.37	−0.12***	0.32	0.33	−0.01
	(0.47)	(0.43)	(0.48)	(0.00)	(0.47)	(0.47)	(0.74)
Observations	1274	282	403	685	685	589	1274

Notes: SOEP v31, waves 2006-2014. Descriptive statistics calculated for our sample based on the youth questionnaire. Standard deviations in parentheses. The fourth and last column depict the difference in means between the mentioned groups and the p-values of a t-test in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 5.2 provides an overview on the descriptive statistics for the sample of 17-year olds, including the set of variables we later add as controls. On average students in our sample felt quite healthy (1.63) and students in the 8 year track describe themselves as slightly more healthy compared to students in the 9 year track (1.55 vs. 1.67). Students' BMI is on average 21 with no significant differences between the two groups of academic track students. When it comes to mental well-being, students on average do not report to worry much (mean=4.4). All three health measures are slightly worse for students in the vocational track schools.

Covariates are quite balanced between 8 year and 9 year group. Slightly more than half of our sample are women, one quarter comes from a non-intact family (meaning not living with both parents for at least one year during childhood) and 19% of children have a parent not born in Germany.[14] Roughly one third have at least one parent who graduated from the academic school track and one third of our students live in rural areas. Our sample of students in the non-academic track has a significantly higher share of non-intact households, migration background and a lower share of parents who graduated from high-school than the sample of academic track students. This reflects general differences in the German student population and is unlikely to have changed during our study period.[15]

[14] Official statistics for the school year 2013/2014 state that 51.9% of lower secondary academic track students and 53.6% of higher secondary academic track students were female (Statistisches Bundesamt, 2014, p. 10).

[15] According to the PISA studies, Germany used to be a country where socio-economic background has been a strong predictor of student performance, recently this relationship weakened and has moved to OECD average (OECD, 2013b, pp. 78ff).

Table 5.3 provides descriptive statistics for our sample used in the post graduation analyses. Self-assessed health of the graduates is slightly worse (2.04) compared to our sample of 17-year old students. BMI is also slightly higher for the student sample and graduates from the academic track after 9 years have a higher BMI compared to students who graduated after 8 years (22.5 and 21.9 respectively). Our graduates have an average MCS of 50.1 which does not significantly differ between subgroups.

Table 5.3: Graduate Sample: Descriptive Statistics

	Pooled	9 Years	8 Years	9 Years - 8 Years	Academic	Vocational	Academic - Vocational
	Mean (S.D.)	Mean (S.D.)	Mean (S.D.)	Difference (P-value)	Mean (S.D.)	Mean (S.D.)	Difference (P-value)
Self-assessed health	2.03	1.95	1.89	0.06	1.93	2.09	−0.15***
	(0.82)	(0.78)	(0.71)	(0.42)	(0.76)	(0.84)	(0.00)
Body mass index	22.91	22.46	21.86	0.60**	22.28	23.22	−0.94***
	(3.46)	(2.97)	(2.71)	(0.04)	(2.91)	(3.67)	(0.00)
MCS	50.09	50.10	49.96	0.14	50.06	50.10	−0.05
	(9.34)	(9.34)	(8.43)	(0.88)	(9.07)	(9.47)	(0.93)
Female	0.50	0.53	0.55	−0.02	0.54	0.48	0.06**
	(0.50)	(0.50)	(0.50)	(0.65)	(0.50)	(0.50)	(0.04)
Age in months	246.49	250.17	237.41	12.76***	246.46	246.50	−0.04
	(8.98)	(7.32)	(7.01)	(0.00)	(9.26)	(8.85)	(0.93)
Non-intact family	0.34	0.20	0.35	−0.15***	0.24	0.39	−0.14***
	(0.47)	(0.40)	(0.48)	(0.00)	(0.43)	(0.49)	(0.00)
Migration background	0.17	0.13	0.11	0.01	0.12	0.19	−0.07***
	(0.37)	(0.33)	(0.32)	(0.68)	(0.33)	(0.39)	(0.00)
High parental education	0.27	0.48	0.51	−0.03	0.49	0.16	0.33***
	(0.44)	(0.50)	(0.50)	(0.60)	(0.50)	(0.36)	(0.00)
Rural	0.34	0.22	0.39	−0.16***	0.27	0.37	−0.10***
	(0.47)	(0.42)	(0.49)	(0.00)	(0.45)	(0.48)	(0.00)
Graduation two years ago	0.48	0.54	0.40	0.15***	0.50	0.47	0.03
	(0.50)	(0.50)	(0.49)	(0.00)	(0.50)	(0.50)	(0.26)
Observations	1387	327	134	461	461	926	1387

Notes: SOEP v31, waves 2008, 2010, 2012 & 2014. Standard deviations in parentheses. The fourth and last column depict the difference in means between the mentioned groups and the p-values of a t-test in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Fewer observations for body mass index and MCS.

The share of women who graduated from academic track (54%) is significantly higher than the share of female graduates from non-academic tracks (48%).[16] Students in the control group were on average 20.8 years old and hence one year older than the reform group which is a mechanic effect of the reform. All other covariates are distributed similarly to the sample of 17-year olds. Graduates from the non-academic track have a higher probability of coming from a non-intact family, having migration background as well as low parental education. There are however significant differences in the covariates in the sample of graduates that we did not observe for the sample of 17-year olds. Graduates from the 8 year track have a higher probability of coming from a non-intact family and having a migration background and living in a rural area. We use a dummy variable to control for the one year difference that occurs as the SOEP includes the health questions only every other year. About 50% of graduates in our sample graduated the year before the survey, the other half graduated one year earlier.

[16] While this difference looks large, it is in line with official statistics for graduates in 2013, where 54.7% of academic track graduates were female (Statistisches Bundesamt, 2014, p. 291).

5.3 Results

In our baseline specification we estimate the model presented above in equation (5.4). We always cluster our standard errors at the year-state-schooltype level to control for within-group error correlation (Angrist and Pischke, 2009). We separately estimate the effects of increased instruction intensity on students in school and after graduation. Then, we analyze how health effects differ if we exclude students in the transition period from the 9 year system to the 8 year system. The last part of this result section includes different robustness checks.

5.3.1 Effects on Students in School

Effects of the increase of instruction intensity on 17 year old students are reported in Table 5.4 where the coefficient of interest is *8 years*. Self-assessed health is not affected by the reform. BMI is higher when instruction intensity is increased but the effect is not significantly different from zero in the pooled sample. We do however find a significant increase by roughly 1.2 BMI points for females.[17] Probit regressions (Table A5.1, in the appendix) show that the higher BMI for women is driven by a higher share of overweight women. A similar pattern can be found for our mental health proxy *worry a lot* which is significantly higher (worse) for females if their instruction intensity was higher. The coefficients remain almost unchanged when we add further covariates to the model. From analyzing these three different health indicators, we conclude that, on average, increased instruction intensity has a negative effect on female students' health – but not for males'.

Table 5.4: Youth Sample: Results

	pooled		**male**		**female**	
	(1)	(2)	(3)	(4)	(5)	(6)
Self-assessed health						
8 years	−0.036	−0.032	−0.022	0.004	−0.049	−0.063
	(0.084)	(0.086)	(0.130)	(0.129)	(0.119)	(0.121)
Additional controls	-	✓	-	✓	-	✓
BMI						
8 years	0.386	0.425	−0.143	−0.056	1.182***	1.167***
	(0.338)	(0.332)	(0.402)	(0.393)	(0.420)	(0.418)
Additional controls	-	✓	-	✓	-	✓
Worry a lot						
8 years	0.380*	0.359	−0.038	−0.046	0.708**	0.709**
	(0.219)	(0.219)	(0.292)	(0.301)	(0.279)	(0.281)
Additional controls	-	✓	-	✓	-	✓
N	1274	1274	624	624	650	650

Notes: SOEP v31 waves 2006-2014. OLS regressions. All estimations include a constant, a maximum set of state and year dummies, an academic-track dummy, interactions of academic-track with state and year dummies, and control for sex in pooled models. Additional controls include age in months, non-intact family, migration background, high parental education, and rural. Standard errors, reported in parentheses, are clustered at wave-state-schooltype level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

[17] For women with a height of 165cm this would mean a difference of 3.3kg.

5.3.2 Effects on Students after Graduation

The SOEP provides a detailed set of health variables only every other year. In order to estimate the health effects of the reform after graduation, we therefore use the earliest health information available starting from the year after a student has graduated. As we control for years since graduation, our results are not driven by the fact that we observe the graduates at different points in time. Table 5.5 summarizes the health effects of the reform after students graduated. For none of our health measures the reform coefficient differs significantly from zero and the coefficients are barely changed by adding additional covariates. Students in the eight year track do not report a different self assessed health and we observe only a slight (insignificant) decrease in BMI and a slight (insignificant) increase in mental health measured by MCS.

Table 5.5: Graduate Sample: Results

	pooled		**male**		**female**	
	(1)	(2)	(3)	(4)	(5)	(6)
Self-assessed health						
8 years	−0.038	−0.027	−0.053	−0.014	−0.035	−0.064
	(0.116)	(0.119)	(0.178)	(0.174)	(0.177)	(0.176)
Additional controls	-	✓	-	✓	-	✓
BMI						
8 years	−0.162	−0.112	−0.198	−0.057	−0.300	−0.400
	(0.515)	(0.492)	(0.554)	(0.535)	(0.704)	(0.674)
Additional controls	-	✓	-	✓	-	✓
MCS						
8 years	1.006	0.896	−0.053	−0.172	2.204	2.453
	(1.282)	(1.304)	(1.720)	(1.585)	(2.571)	(2.665)
Additional controls	-	✓	-	✓	-	✓
N	1387	1387	699	699	688	688

Notes: SOEP v31 waves 2008, 2010, 2012 & 2014. OLS regressions. All estimations include a constant, a maximum set of state and year dummies, an academic-track dummy, interactions of academic-track with state and year dummies, a dummy indicating whether the student graduated two years ago, age in months, and gender (only in pooled models). Additional controls include non-intact family, migration background, high parental education, and rural. Standard errors, reported in parentheses, are clustered at the wave-state-schooltype level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

After graduation we do not anymore observe the negative health effects the reform has on females in school. This is in line with the results from Forbes et al. (2017) who show that longer schooling can have negative mental health effects on females. If women profit from having to spend less years in school, this can offset negative health effects during school. A potential mechanism is that students affected by the reform have the freedom to chose their life paths after school at a younger age. This freedom can explain the disappearance of the negative effect estimated for 17-year olds if subsequent life choices improve mental well-being. Findings by Meyer and Thomsen (2016) support this explanation. Students affected by the reform have a higher probability of doing an internship or spending a year abroad – probably relieving them from the extra pressure of increased instruction intensity in school.

5.3.3 Double Cohort as a Moderator

By shortening the academic track, each state generated a so called *double cohort*, i. e. students from the last nine year as well as from the first eight year scheme graduated together. Although universities did prepare for a higher number of students, the perceived higher competition for places at universities and apprenticeship positions could have induced additional stress. Additionally, the reform was implemented before all details were agreed on. This led to large uncertainties for students in the first reform cohort. As our next analysis step, we therefore exclude students from the *double cohort* from our sample.

Table 5.6: Youth Sample: Excluding the Double Cohorts

	pooled		**male**		**female**	
	(1)	(2)	(3)	(4)	(5)	(6)
Self-assessed health						
8 years	0.063	0.063	0.054	0.084	0.126	0.096
	(0.132)	(0.137)	(0.159)	(0.158)	(0.198)	(0.211)
Additional controls	-	✓	-	✓	-	✓
BMI						
8 years	0.659	0.643	−0.170	−0.203	1.130*	1.093*
	(0.446)	(0.441)	(0.593)	(0.574)	(0.625)	(0.637)
Additional controls	-	✓	-	✓	-	✓
Worry a lot						
8 years	0.541**	0.499*	−0.295	−0.326	1.119***	1.102***
	(0.257)	(0.267)	(0.495)	(0.486)	(0.371)	(0.407)
Additional controls	-	✓	-	✓	-	✓
N	1054	1054	500	500	554	554

Notes: SOEP v31 waves 2006-2014. OLS regressions. All estimations include a constant, a maximum set of state and year dummies, an academic-track dummy, interactions of academic-track with state and year dummies, and control for sex in pooled models. Additional controls include age in months, non-intact family, migration background, high parental education, and rural. Standard errors, reported in parentheses, are clustered at wave-state-schooltype level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Compared to the full sample, this exclusion leads to similar coefficients (Table 5.6). Overall we conclude that the general pattern of negative health effects of increased instruction intensity on females in school is not driven by the special circumstances of the *double cohort*.

If we repeat our baseline analysis for graduates without those in the *double cohorts*, we find that the slight positive health effects observed for the full sample increase and become statistically significantly different from zero for females. Women with 8 years of secondary school have a 1.7 to 2.4 point lower BMI after graduation (driven by a decreased risk of being overweight) and have a MCS that is increased by between 50 to 60% of a standard deviation. We therefore conclude that the stress of being in a *double cohort* cancels out positive health effects of the reform after students left school. This positive effect can have at least three reasons. First, as noted above, students have more flexibility in choosing their life paths, enabling them to make life choices that benefit their health one year earlier. Second, increased instruction intensity compressed the time during which students acquired health relevant abilities. This might have enabled them to start living a healthier life earlier. A third explanation is that students who experienced increased instruction intensity learned to cope with stress during school and are now better able to deal with difficult life situations after graduation.

Table 5.7: Graduate Sample: Excluding the Double Cohorts

	pooled		male		female	
	(1)	(2)	(3)	(4)	(5)	(6)
Self-assessed health						
8 years	−0.061	−0.068	0.086	0.128	−0.234	−0.288
	(0.115)	(0.129)	(0.219)	(0.238)	(0.207)	(0.225)
Additional controls	-	✓	-	✓	-	✓
BMI						
8 years	−0.373	−0.404	1.092	1.022	−2.222**	−2.201**
	(0.462)	(0.418)	(0.976)	(0.988)	(0.947)	(0.877)
Additional controls	-	✓	-	✓	-	✓
MCS						
8 years	1.990	2.114	−1.558	−1.681	6.137**	6.564**
	(2.091)	(2.158)	(2.930)	(2.902)	(2.847)	(3.072)
Additional controls	-	✓	-	✓	-	✓
N	1097	1097	549	549	548	548

Notes: SOEP v31 waves 2008, 2010, 2012 & 2014. OLS regressions. All estimations include a constant, a maximum set of state and year dummies, an academic-track dummy, interactions of academic-track with state and year dummies, and control for sex (only in pooled models), Graduation two years ago, and age in months. Additional controls include non-intact family, migration background, high parental education, and rural. Standard errors, reported in parentheses, are clustered at wave-state-schooltype level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

5.3.4 Robustness Checks

We conduct several robustness checks, changing our sample composition, employing maximum likelihood estimation techniques and adjusting standard errors.

Similar time to examinations

Time to graduation might influence students' health status as the final exams are a major event that determines which universities graduates can go to or which subjects they can study. As a matter of fact, students in the 8 year track are on average closer to their final examinations than the control group. We show that restricting the sample to students with similar time until graduation does give qualitatively similar results (Table A5.5, in the Appendix).

Heterogeneity by parental education

To assess potential social inequalities in the burden of the reform we conducted additional subgroup analyses by parental education. The results for the youth sample (Table A5.8, in the Appendix) indicate that the effect is rather driven by children with at least one parent who finished higher secondary education. The results for the graduate sample (Table A5.9, in the Appendix) are less straightforward to interpret, it seems as if again there are differences between students of different backgrounds, but no clear direction can be told. These heterogeneity checks are based on a comparably sample of 173 observations and might therefore not be too reliable overall.

Excluding states with few observations

In some states we observe only very few individuals in the treatment or control group. The fewer students we observe per state, the higher is the probability of drawing students who are not at all representative of the state's student body. Therefore, we also restricted our sample to contain only states, where we observe at least 10 students in each state's treatment and control groups. This leaves us with seven states in the sample of 17-year old students (see Table A5.6, in the Appendix) and five states in the sample of graduates (Table A5.7, in the Appendix). The results for the students who are still in school are similar to our baseline results. The only notable difference is that now the effect on worrying a lot is also significantly positive in the pooled estimation. When we additionally exclude the double cohort (Table A5.6, second row) the results are very similar to our baseline specification. In the sample of graduates we again see no significant health effects of the reform in the full sample yet again find positive health effects of the reform once removing the double cohort.

Only academic track students

Coefficients for increased instruction intensity from our triple difference in differences estimations are identical to simple difference in differences estimations when no additional controls are included (see equation 5.3). To make sure that our results are not driven by changes in control variables for the non academic track students we also estimate simple difference-in-differences models . Results for our sample of 17-year old students as well as for graduates are very similar to our baseline specification (Table A5.3 and A5.4, in the Appendix).

Probit regressions

We also run probit and ordered probit regressions for the binary or ordinal dependent variables. We find qualitatively very similar results for our student sample as well as for our sample of graduates (Tables A5.1 and A5.2, in the Appendix).

Standard Error Corrections

As Bertrand et al. (2004) point out, standard errors in difference-in-differences settings can be biased downwards due to serial or within cluster correlation. We address this concern in three ways. First by changing the cluster-level, second by bootstrapping the standard errors and third by running placebo tests.

We decided to cluster standard errors from all previous estimations at the year-state-schooltype level because this is the unit where the variation is coming from (Angrist and Pischke, 2009). One could however argue that when it comes to the German school system, differences between federal states are more important than differences over time and hence clustering must be on the state level. If we reestimate the models from above and cluster at the state level, standard errors marginally increase but significance levels are not affected. Following Cameron et al. (2011), we also apply wild clustered bootstrapping for estimates clustered at the state level.[18] This again marginally increases standard errors but does not change the overall picture (Tables A5.10 and A5.11).[19]

Finally, we conduct placebo tests suggested by Chetty et al. (2009). We run 2000 placebo regressions where we randomly assign the introduction of the 8 year academic track in the states 100 times and for each of these draws randomly assign for each student whether she is in the

---

[18] Bootstrapping does not change standard errors from our baseline analysis because there are always more than 50 clusters.

[19] We used the Caskey (2015) `cgmwildboot.ado` for bootstrapping.

academic track 20 times. We then compare the distribution of coefficients from the placebo regressions with the actual coefficients. The results displayed in Figure A5.1 (in the Appendix) show that the coefficients for *BMI* and *worrying* in our full sample of 17-year old females are higher than 95% of the placebo coefficients while the other coefficients are comparable to average coefficients from random treatment assignment. This is also the case for the worrying coefficient if we exclude the double cohort, indicating that the results in Tables 5.4 and 5.6 are not driven by downward biased standard errors. The same holds for our sample of graduates. Our placebo test (Figure A5.2, in the Appendix) indicates that the significant reduction in women's BMI and increase in women's MCS in Tables 5.5 and 5.7 is not due to downward biased standard errors.

5.4 Conclusion

A large literature examines the effect of education on health using changes in compulsory schooling laws to instrument *increases* in years of schooling. It is however unclear how years of schooling causally affect health. We contribute to this literature by examining a reform in Germany where the academic track of secondary education was *reduced* from nine to eight years without changes in total hours taught. The sequential introduction of this reform in different federal states enables us to employ a triple difference-in-differences strategy to estimate the effect of increased instruction intensity on students' health in school and additionally the effect of leaving school one year earlier. We conduct our analysis with data from the German Socio-Economic Panel.

We find worse health status for females who experienced increased instruction intensity – but only as long as they are in school. 17-year old women affected by the reform worry more and also have a higher body mass index. Our results are robust to various sample restrictions and different estimator choices. This suggests that higher instruction intensity in schools translates into worse health outcomes.

Even though the health effects of the reform are negative when females are in school, the effects disappear after graduation and even turn positive once students from the transition period are excluded from our sample. Further research is needed to distinguish between three likely explanations for this effect. Increased instruction intensity could lead to health skills formation being effective earlier in life, leaving more time to positively influence healthy lifestyles. Also, the fact that students can influence their life path more directly one year earlier leads to life choices that benefit health. A third explanation would be that students who experience higher instruction intensity get used to stress during adolescence and are more able to deal with difficult situations later on. Although our analysis cannot identify the mechanism between reduced years with increased instruction intensity and health, our results shed a critical light on the discussion of whether the health effects of education should be mainly analyzed in terms of marginal increases of school years.

It is important to note that in our study a selective group of people is affected. Increases in mandatory school years usually affect a whole cohort while the reform we study only affects academic track students. Although this group is large (about one third of German students graduate from academic track schools), it is in general a self selection of higher ability students. It is therefore possible that a similar reform on lower ability vocational track students would result in different health effects.

Also, a puzzling result from our analysis is that males seem not to be affected by increased instruction intensity at all. They neither exhibit negative health effects in school nor do they profit from positive health effects after graduation. More research on the channels through which instruction intensity affects health might help explain this puzzle.

5.5 Appendix

Table A5.1: Youth Sample: Probit and Ordered Probit Estimations

	Self-assessed health			Overweight (0/1)			Worry a lot		
	(1) pooled	(2) male	(3) female	(4) pooled	(5) male	(6) female	(7) pooled	(8) male	(9) female
Full Sample									
8 years	−0.046	−0.043	−0.076	0.424*	−0.119	1.304***	0.241*	0.003	0.440**
	(0.135)	(0.232)	(0.181)	(0.227)	(0.306)	(0.482)	(0.133)	(0.173)	(0.171)
N	1274	624	650	1255	574	586	1274	624	650
No Double Cohort									
8 years	0.114	0.127	0.190	0.753**	−0.061	1.914***	0.350**	−0.169	0.735***
	(0.235)	(0.307)	(0.350)	(0.362)	(0.524)	(0.572)	(0.161)	(0.321)	(0.227)
N	1054	500	554	1036	453	498	1054	500	554

Notes: SOEP v31 waves 2006-2014. Ordered Probit estimations. All estimations include a maximum set of state and year dummies, an academic-track dummy, and interactions of academic-track with state and year dummies. Standard errors, reported in parentheses, are clustered at wave-state-schooltype level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A5.2: Graduate Sample: Probit and Ordered Probit Estimations

	Self-assessed health			Overweight (0/1)		
	(1) pooled	(2) male	(3) female	(4) pooled	(5) male	(6) female
Full sample						
8 years	−0.043	−0.072	−0.028	−0.317	−0.153	−0.537**
	(0.166)	(0.252)	(0.237)	(0.238)	(0.397)	(0.238)
N	1387	699	688	1387	699	682
No double cohort						
8 years	−0.041	0.197	−0.289	−0.474	0.272	−1.676***
	(0.164)	(0.342)	(0.283)	(0.386)	(0.658)	(0.431)
N	1097	549	548	1095	549	540

Notes: SOEP v31 waves 2008, 2010, 2012 & 2014. Ordered Probit estimation. All estimations include a maximum set of state and year dummies, an academic-track dummy, and interactions of academic-track with state and year dummies. Standard errors, reported in parentheses, are clustered at wave-state-schooltype level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A5.3: Youth Sample: Only Academic Track Students

	pooled		male		female	
	(1)	(2)	(3)	(4)	(5)	(6)
Self-assessed health						
8 years	−0.034	−0.037	−0.022	−0.022	−0.049	−0.057
	(0.083)	(0.083)	(0.131)	(0.134)	(0.119)	(0.120)
Additional controls	-	✓	-	✓	-	✓
BMI						
8 years	0.383	0.420	−0.143	−0.094	1.182***	1.149***
	(0.342)	(0.337)	(0.403)	(0.403)	(0.421)	(0.423)
Additional controls	-	✓	-	✓	-	✓
Worry a lot						
8 years	0.376*	0.373*	−0.038	−0.020	0.708**	0.739**
	(0.214)	(0.211)	(0.293)	(0.302)	(0.279)	(0.284)
Additional controls	-	✓	-	✓	-	✓
N	685	685	331	331	354	354

Notes: SOEP v31 waves 2006-2014. OLS regressions. All estimations include a maximum set of state and year dummies and control for sex (only in pooled models). Additional controls include age in months, non-intact family, migration background, high parental education, and rural. Standard errors, reported in parentheses, are clustered at wave-state level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A5.4: Graduate Sample: Only Academic Track Students

	pooled		male		female	
	(1)	(2)	(3)	(4)	(5)	(6)
Self-assessed health						
8 years	0.026	0.017	0.106	0.134	−0.050	−0.070
	(0.143)	(0.134)	(0.173)	(0.149)	(0.251)	(0.248)
Additional controls	-	✓	-	✓	-	✓
BMI						
8 years	−0.568	−0.483	−1.102	−0.853	−0.360	−0.328
	(0.632)	(0.588)	(0.891)	(0.808)	(0.705)	(0.745)
Additional controls	-	✓	-	✓	-	✓
MCS						
8 years	−0.167	−0.060	−1.638	−1.691	1.547	2.055
	(1.720)	(1.704)	(2.333)	(2.148)	(2.988)	(2.941)
Additional controls	-	✓	-	✓	-	✓
N	461	461	214	214	247	247

Notes: SOEP v31 waves 2008, 2010, 2012 & 2014. OLS regressions. All estimations include a maximum set of state and year dummies. Standard errors, reported in parentheses, are clustered at wave-state level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A5.5: Youth Sample: Only Similar Time to Examinations

	Self-assessed health			BMI			Worry a lot		
	(1) pooled	(2) male	(3) female	(4) pooled	(5) male	(6) female	(7) pooled	(8) male	(9) female
Similar time to grad.									
8 years	−0.090	0.036	−0.192	0.073	−0.674*	1.141***	0.295	−0.047	0.606*
	(0.088)	(0.134)	(0.132)	(0.311)	(0.405)	(0.417)	(0.257)	(0.342)	(0.325)
N	946	448	498	946	448	498	946	448	498
R^2	0.048	0.087	0.100	0.069	0.082	0.098	0.079	0.070	0.081
No double cohort									
8 years	−0.096	0.029	−0.207	0.493	−0.778	1.449**	0.518*	−0.210	1.186***
	(0.130)	(0.159)	(0.188)	(0.480)	(0.709)	(0.620)	(0.285)	(0.483)	(0.446)
N	791	363	428	791	363	428	791	363	428
R^2	0.056	0.099	0.119	0.088	0.087	0.127	0.084	0.079	0.091

Notes: SOEP v31 waves 2006-2014. OLS regressions. All estimations include a constant, a maximum set of state and year dummies, an academic-track dummy, interactions of academic-track with state and year dummies, and control for sex in pooled estimations. Standard errors, reported in parentheses, are clustered at wave-state-schooltype level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A5.6: Youth Sample: Only States with many Observations

	Self-assessed health			BMI			Worry a lot		
	(1) pooled	(2) male	(3) female	(4) pooled	(5) male	(6) female	(7) pooled	(8) male	(9) female
All									
8 years	−0.049	−0.038	−0.040	0.249	−0.133	1.029**	0.618***	0.162	1.014***
	(0.102)	(0.147)	(0.155)	(0.381)	(0.454)	(0.493)	(0.220)	(0.299)	(0.321)
N	1129	552	577	1129	552	577	1129	552	577
R^2	0.026	0.055	0.042	0.050	0.042	0.069	0.070	0.045	0.061
No double cohort									
8 years	0.238	0.130	0.351	0.725*	0.304	1.371*	0.897***	0.027	1.672***
	(0.162)	(0.191)	(0.270)	(0.402)	(0.490)	(0.695)	(0.298)	(0.548)	(0.529)
N	939	443	496	939	443	496	939	443	496
R^2	0.031	0.072	0.052	0.067	0.052	0.096	0.073	0.046	0.074

Notes: SOEP v31 waves 2006-2014. This sample contains the following states: Lower Saxony, Baden-Württemberg, Bavaria, North-Rhine-Westphalia, Berlin, Brandenburg and Saxony-Anhalt. OLS regressions. All estimations include a constant, a maximum set of state and year dummies, an academic-track dummy, interactions of academic-track with state and year dummies, and control for sex in pooled estimations. Standard errors, reported in parentheses, are clustered at wave-state-schooltype level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A5.7: Graduate Sample: Only States with many Observations

	Self-assessed health			BMI			MCS		
	(1) pooled	(2) male	(3) female	(4) pooled	(5) male	(6) female	(7) pooled	(8) male	(9) female
All									
8 years	−0.019	−0.139	0.125	−0.550	−0.354	−0.622	0.686	0.824	1.335
	(0.128)	(0.195)	(0.172)	(0.482)	(0.633)	(0.594)	(0.949)	(1.765)	(1.799)
N	1188	597	591	1188	597	591	1188	597	591
R^2	0.040	0.046	0.048	0.093	0.055	0.058	0.060	0.056	0.075
No double cohort									
8 years	−0.139	−0.131	−0.268	−1.270***	0.383	−3.029***	3.524	0.550	9.928**
	(0.115)	(0.246)	(0.168)	(0.352)	(0.986)	(0.955)	(2.805)	(3.444)	(3.867)
N	944	469	475	944	469	475	944	469	475
R^2	0.043	0.051	0.064	0.106	0.070	0.075	0.066	0.055	0.106

Notes: SOEP v31 waves 2008, 2010, 2012 & 2014. This sample contains the following states: Lower Saxony, Baden-Württemberg, Bavaria, North-Rhine-Westphalia, and Saxony-Anhalt. OLS regressions. All estimations include a constant, a maximum set of state and year dummies, an academic-track dummy, interactions of academic track with state and year dummies, and control for Abitur two years ago, age in months and in pooled estimations also for sex. Standard errors, reported in parentheses, are clustered at wave-state-schooltype level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A5.8: Youth Sample: Heterogeneity by Parental Education

	Low parental education			High parental education		
	(1) pooled	(2) male	(3) female	(4) pooled	(5) male	(6) female
Self-assessed health						
8 years	−0.067	−0.107	0.001	0.006	0.108	−0.198
	(0.125)	(0.181)	(0.163)	(0.123)	(0.178)	(0.194)
BMI						
8 years	−0.050	−0.806	0.962	0.723*	0.298	0.950*
	(0.495)	(0.625)	(0.694)	(0.391)	(0.572)	(0.526)
Worry a lot						
8 years	0.065	0.012	0.151	0.648**	0.099	0.998**
	(0.324)	(0.557)	(0.454)	(0.265)	(0.288)	(0.492)
N	819	393	426	442	224	218

Notes: SOEP v31 waves 2006-2014. OLS regressions. All estimations include a constant, a maximum set of state and year dummies, an academic-track dummy, interactions of academic-track with state and year dummies, and control for sex in pooled models. Standard errors, reported in parentheses, are clustered at wave-state-schooltype level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A5.9: Graduate Sample: Heterogeneity by Parental Education

	Low parental education			High parental education		
	(1) pooled	(2) male	(3) female	(4) pooled	(5) male	(6) female
Self-assessed health						
8 years	−0.030	−0.229	0.174	−0.299	−0.237	−0.276
	(0.149)	(0.256)	(0.340)	(0.240)	(0.345)	(0.353)
BMI						
8 years	−1.216**	−1.684	−0.248	1.104	2.004***	0.100
	(0.553)	(1.111)	(0.564)	(0.719)	(0.629)	(1.176)
MCS						
8 years	1.288	−0.108	4.090	2.990	2.423	1.570
	(1.894)	(2.159)	(5.523)	(2.149)	(4.538)	(3.430)
N	877	457	420	357	173	184

Notes: SOEP v31 waves 2008, 2010, 2012 & 2014. OLS regressions. All estimations include a constant, a maximum set of state and year dummies, an academic-track dummy, interactions of academic-track with state and year dummies, and control for sex (only in pooled models), Graduation two years ago, and age in months. Standard errors, reported in parentheses, are clustered at wave-state-schooltype level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A5.10: Youth Sample: Standard error calculations

			p-values from clustering		
		Coefficient	State-year-school type	State	State bootstrapped
Self-assessed health					
	pooled	−0.036	0.670	0.662	0.654
	male	−0.022	0.868	0.817	0.812
	female	−0.049	0.683	0.603	0.720
BMI					
	pooled	0.386	0.255	0.295	0.308
	male	−0.143	0.723	0.739	0.728
	female	1.182	0.006	0.019	0.030
Worry					
	pooled	0.380	0.085	0.175	0.202
	male	−0.038	0.896	0.867	0.858
	female	0.708	0.012	0.074	0.176

Notes: SOEP v31 waves 2006-2014. Coefficients on the reform dummy from OLS regressions. All estimations include a constant, a maximum set of state and wave dummies, an academic track dummy, interactions of academic track with state and wave dummies and control for gender in pooled models (reference category: male). P-values are obtained from conventionally clustered standard errors at wave-state-schooltype level and at state level in the first two columns. The third column displays p-values obtained from wild-cluster bootstrapping at state level.

Table A5.11: Graduate Sample: Standard error calculations

			p-values from clustering		
Self-assessed health		Coefficient	State-year-school type	State	State bootstrapped
	pooled	−0.038	0.747	0.819	0.836
	male	−0.053	0.766	0.817	0.770
	female	−0.035	0.842	0.870	0.874
BMI					
	pooled	−0.162	0.755	0.778	0.868
	male	−0.198	0.722	0.720	0.692
	female	−0.300	0.671	0.715	0.764
MCS					
	pooled	1.006	0.435	0.471	0.464
	male	−0.053	0.976	0.979	0.986
	female	2.204	0.394	0.383	0.364

Notes: SOEP v31 waves 2008, 2010, 2012 & 2014. Coefficients on the reform dummy from OLS regressions. All estimations include a constant, a maximum set of state and wave dummies, an academic track dummy, interactions of academic track with state and wave dummies and control for graduation two years ago, age, and additionally for gender in pooled models (reference category: male). P-values are obtained from conventionally clustered standard errors at wave-state-schooltype level and at state level in the first two columns. The third column displays p-values obtained from wild-cluster bootstrapping at state level.

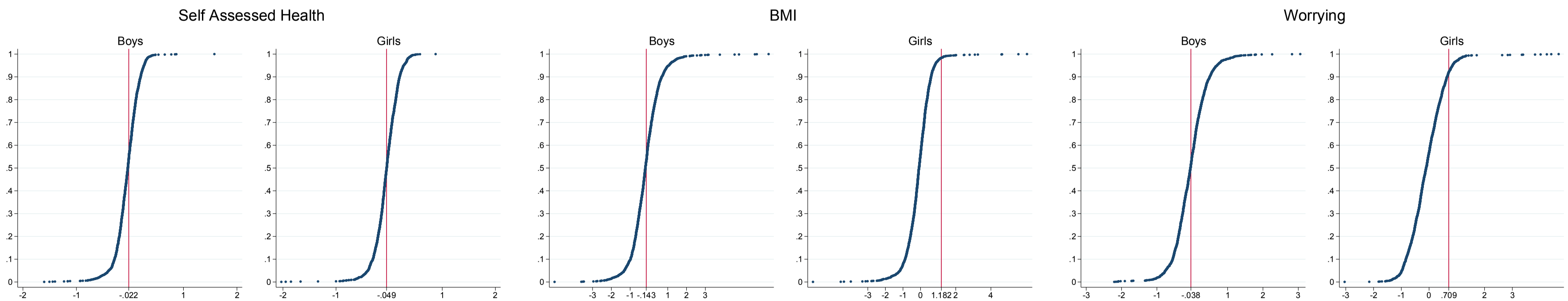
Table A5.12: Youth Sample (excluding double cohort): Standard error calculations

			p-values from clustering		
Self-assessed health		Coefficient	State-year-school type	State	State bootstrapped
	pooled	0.063	0.636	0.663	0.710
	male	0.054	0.736	0.662	0.674
	female	0.126	0.528	0.451	0.526
BMI					
	pooled	0.659	0.141	0.224	0.254
	male	−0.170	0.774	0.839	0.876
	female	1.130	0.073	0.056	0.068
Worry					
	pooled	0.541	0.037	0.046	0.038
	male	−0.295	0.552	0.397	0.372
	female	1.119	0.003	0.024	0.054

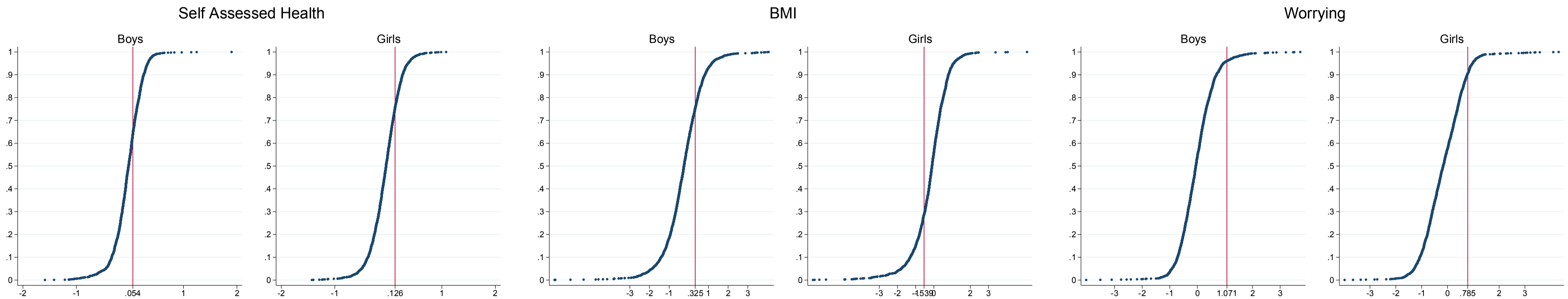
Notes: SOEP v31 waves 2006-2014. Coefficients on the reform dummy from OLS regressions. All estimations include a constant, a maximum set of state and wave dummies, an academic track dummy, interactions of academic track with state and wave dummies and control for gender in pooled models (reference category: male). P-values are obtained from conventionally clustered standard errors at wave-state-schooltype level and at state level in the first two columns. The third column displays p-values obtained from wild-cluster bootstrapping at state level.

Figure A5.1: Youth Sample: Distribution of Coefficients from Placebo Regressions

All Observations



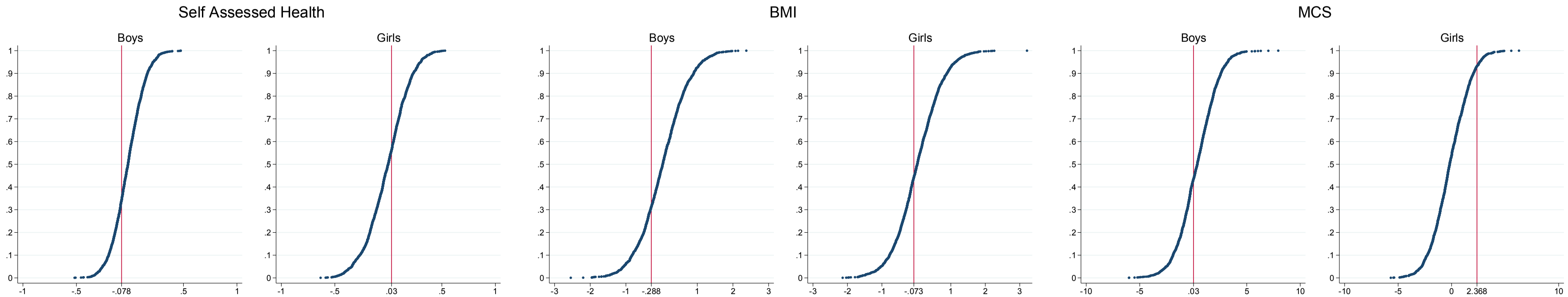
Excluding (Placebo) Double Cohort



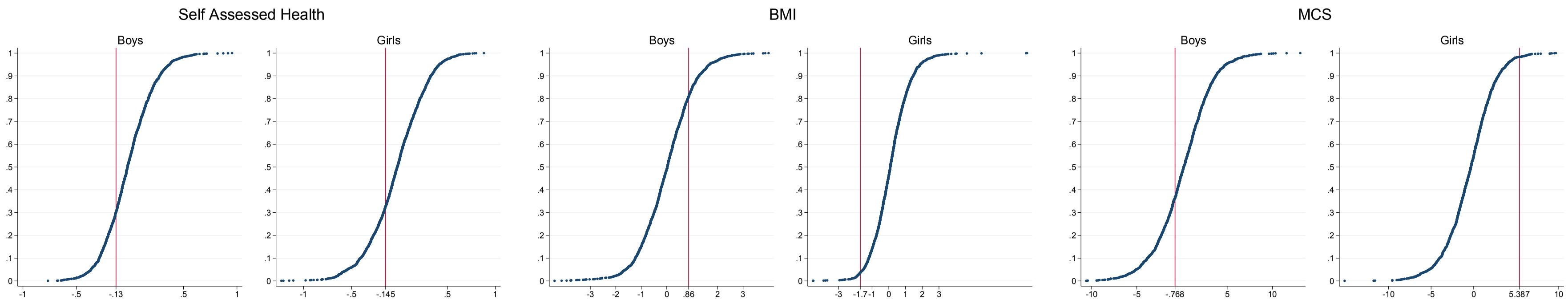
Note: Each figure plots the empirical distribution of regression coefficients from 2000 placebo regressions with random treatment assignment. Vertical lines indicate the treatment effect from Table 5.4.

Figure A5.2: Graduate Sample: Distribution of Coefficients from Placebo Regressions

All Observations



Excluding (Placebo) Double Cohort



Note: Each figure plots the empirical distribution of regression coefficients from 2000 placebo regressions with random treatment assignment. Vertical lines indicate the treatment effect from Table 5.5.

Table A5.13: Graduate Sample (excluding double cohort): Standard error calculations

			p-values from clustering		
		Coefficient	State-year-school type	State	State bootstrapped
Self-assessed health					
	pooled	−0.061	0.597	0.743	0.780
	male	0.086	0.697	0.791	0.854
	female	−0.234	0.262	0.295	0.262
BMI					
	pooled	−0.373	0.422	0.448	0.408
	male	1.092	0.267	0.283	0.284
	female	−2.222	0.022	0.067	0.148
MCS					
	pooled	1.990	0.345	0.447	0.442
	male	−1.558	0.597	0.683	0.664
	female	6.137	0.035	0.079	0.146

Notes: SOEP v31 waves 2008, 2010, 2012 & 2014. Coefficients on the reform dummy from OLS regressions. All estimations include a constant, a maximum set of state and wave dummies, an academic track dummy, interactions of academic track with state and wave dummies and control for graduation two years ago, age, and additionally for gender in pooled models (reference category: male). P-values are obtained from conventionally clustered standard errors at wave-state-schooltype level and at state level in the first two columns. The third column displays p-values obtained from wild-cluster bootstrapping at state level.

Table A5.14: Description of Variables

Variable	**Description**
Self-assessed health	Self-reported health on a scale from (1) very good to (5) bad.
Body mass index (BMI)	calculated from self-reported weight and height.
Overweight	Dummy indicating whether BMI $>=$ 25.
Worry a lot	"I see myself as someone who worries a lot." On a scale from (1) not at all to (7) very much.
Mental component scale (MCS)	standardized compound measure of mental well-being normalized to mean 50 and standard deviation 10.
8years	Dummy indicating whether student affected by the reform.
Female	Dummy indicating whether student is female.
Age (months)	Student age in months.
Migration background	Dummy indicating whether student was has at least one parent who was not born in Germany.
Rural	Dummy indicating whether student lives in a rural area.
Non-intact family	Dummy indicating whether student comes from a non-intact family i. e. has not lived in one household with both parents for at least one year before the youth survey.
High parental education	Dummy indicating whether at least one parent finished higher secondary education.
Academic Track	Dummy indicating whether students visited the academic track (*Gymnasium*).
Abitur two years ago	Dummy indicating whether student graduated (hypothetically for vocational tracks) two years prior to the survey.
Double cohort	Dummy indicating whether student graduated in a double cohort.

Bibliography

Akerlund, D., B. H. H. Golsteyn, H. Gronqvist, and L. Lindahl (2016). Time discounting and criminal behavior, *Proceedings of the National Academy of Sciences of the United States of America*, **113** (22), pp. 6160–6165.

Albouy, V. and L. Lequien (2009). Does compulsory education lower mortality? *Journal of Health Economics*, **28** (1), pp. 155–168.

Altner, H. (2007). Naturwissenschaften im achtjährigen Gymnasium G8, *Der mathematische und naturwissenschaftliche Unterricht*, **60** (8), pp. 493–497.

Andersen, H. H., A. Mühlbacher, M. Nübling, J. Schupp, and G. G. Wagner (2007). Computation of standard values for physical and mental health scale scores using the SOEP version of SF-12v2, *Schmollers Jahrbuch*, **127** (1), pp. 171–182.

Andersen, S. C., M. K. Humlum, and A. B. Nandrup (2016). Increasing instruction time in school does increase learning, *Proceedings of the National Academy of Sciences*, **113** (27), pp. 7481–7484.

Andersen, S., G. W. Harrison, M. I. Lau, and E. E. Rutström (2008). Eliciting risk and time preferences, *Econometrica*, **76** (3), pp. 583–618.

Angrist, J. D. and A. B. Krueger (1991). Does compulsory school attendance affect schooling and earnings? *Quarterly Journal of Economics*, **106** (4), pp. 979–1014.

Angrist, J. D. and J.-S. Pischke (2009). *Mostly Harmless Econometrics - An Empiricist's Companion*: Princeton University Press.

Arendt, J. N. (2005). Does education cause better health? A panel data analysis using school reforms for identification, *Economics of Education Review*, **24** (2), pp. 149–160.

Autorengruppe Bildungsberichterstattung (2010). *Bildung in Deutschland 2010. Ein indikatorengestützter Bericht mit einer Analyse zu Perspektiven des Bildungswesens im demografischen Wandel.*, Bielefeld: W. Bertelsmann.

Avendanoy, M., A. de Coulonz, and V. Nafilyan (2017). Does more education always improve mental health? Evidence from a British compulsory schooling reform, *Health, Econometrics and Data Group Discussion Paper 17/10.*

Balart, P., M. Oosterveen, and D. Webbink (2018). Test scores, noncognitive skills and economic growth, *Economics of Education Review*, **63**, pp. 134 – 153.

Barro, R. J. (2001). Human Capital and Growth, *American Economic Review*, **91** (2), pp. 12–17.

Bartling, B., E. Fehr, B. Fischer, F. Kosse, M. Maréchal, F. Pfeiffer, D. Schunk, J. Schupp, C. K. Spieß, and G. G. Wagner (2010). Determinanten kindlicher Geduld – Ergebnisse einer Experimentalstudie im Haushaltskontext, *Schmollers Jahrbuch*, **130** (3), pp. 297–323.

Bauer, P. and R. T. Riphahn (2006). Timing of school tracking as a determinant of intergenerational transmission of education, *Economics Letters*, **91** (1), pp. 90–97.

Becker, A., T. Deckers, T. Dohmen, A. Falk, and F. Kosse (2012). The Relationship Between Economic Preferences and Psychological Personality Measures, *Annual Review of Economics*, **4** (1), pp. 453–478.

Becker, G. S. (1962). Investment in Human Capital: A Theoretical Analysis, *Journal of Political Economy*, **70** (5, Part 2), pp. 9–49.

——— (1993). *Human Capital: A Theoretical and Empirical Analysis, with Special Reference to Education*, Chicago: The University of Chicago Press, 3rd edition.

Ben-Porath, Y. (1967). The Production of Human Capital and the Life Cycle of Earnings, *Journal of Political Economy*, **75** (4, Part 1), pp. 352–365.

Benjamin, D. J., S. A. Brown, and J. M. Shapiro (2013). Who is "behavioral"? Cognitive ability and anomalous preferences, *Journal of the European Economic Association*, **11** (6), pp. 1231–1255.

Bergmüller, S. (2007). *Schulstress unter Jugendlichen: Entstehungsbedingungen, vermittelnde Prozesse und Folgen ; eine empirische Studie im Rahmen von PISA 2003*, **Bd. 24**, Schriftenreihe Studien zur Streßforschung, Hamburg: Kovač.

Bertrand, M., E. Duflo, and S. Mullainathan (2004). How much should we trust differences-in-differences estimates? *Quarterly Journal of Economics*, **119** (1), pp. 249–275.

Bettinger, E. and R. Slonim (2007). Patience among children, *Journal of Public Economics*, **91** (1-2), pp. 343–363.

Bickel, W. K., A. L. Odum, and G. J. Madden (1999). Impulsivity and cigarette smoking: Delay discounting in current, never, and ex-smokers, *Psychopharmacology*, **146** (4), pp. 447–454.

Bilz, L. (2008). *Schule und psychische Gesundheit: Risikobedingungen für emotionale Auffälligkeiten von Schülerinnen und Schülern*, Wiesbaden: VS Verlag für Sozialwissenschaften.

Bilz, L., C. Hähne, and W. Melzer (2003). Die Lebenswelt Schule und ihre Auswirkungen auf die Gesundheit von Jugendlichen, in K. Hurrelmann, A. Klocke, W. Melzer, and U. Ravens-Sieberer (Eds.) *Jugendgesundheitssurvey*, Weinheim: Juventa, pp. 243–299.

Blossfeld, H.-P., H.-G. Roßbach, and J. v. Maurice (Eds.) (2011). *Education as a lifelong process: The German National Educational Panel Study (NEPS) Zeitschrift für Erziehungswissenschaft, 14*, Wiesbaden: VS Verlag für Sozialwissenschaften.

Bochner, S. and K. H. David (1968). Delay of gratification, age and intelligence in an Aboriginal culture, *International Journal of Psychology*, **3** (3), pp. 167–174.

Böhm-Kasper, O. and H. Weishaupt (2002). Belastung und Beanspruchung von Lehrern und Schülern am Gymnasium, *Zeitschrift für Erziehungswissenschaft*, **5** (3), pp. 472–499.

Bradford, W. D. (2010). The Association Between Individual Time Preferences and Health Maintenance Habits, *Medical Decision Making*, **30** (1), pp. 99–112.

Brunello, G. and D. Checchi (2007). Does school tracking affect equality of opportunity? New international evidence, *Economic policy*, **22** (52), pp. 782–861.

Brunello, G., D. Fabbri, and M. Fort (2013). The causal effect of education on body mass: Evidence from Europe, *Journal of Labor Economics*, **31** (1), pp. 195–223.

Brunello, G., M. Fort, and G. Weber (2009). Changes in compulsory schooling, education and the distribution of wages in Europe, *Economic Journal*, **119** (536), pp. 516–539.

Brunello, G. and M. Schlotter (2011). Non Cognitive Skills and Personality Traits: Labour Market Relevance and their Development in Education & Training Systems, IZA Discussion Papers 5743, Institute for the Study of Labor (IZA).

Brütting, R. (2007). Drastische Kürzung des Geschichtsunterrichts im achtjährigen Gymnasium, *Geschichte, Politik und ihre Didaktik*, **35** (3-4), pp. 179–182.

Bub, K. L., L. E. Robinson, and D. S. Curtis (2016). Longitudinal associations between self-regulation and health across childhood and adolescence, *Health Psychology*, **35** (11), pp. 1235–1245.

Büttner, B. and S. L. Thomsen (2015). Are We Spending Too Many Years in School? Causal Evidence of the Impact of Shortening Secondary School Duration, *German Economic Review*, **16** (1), pp. 65–86.

Cadena, B. C. and B. J. Keys (2015). Human capital and the lifetime costs of impatience, *American economic journal : a journal of the American Economic Association*, **7** (3), pp. 126–153.

Cameron, A. C., J. B. Gelbach, and D. L. Miller (2011). Robust inference with multiway clustering, *Journal of Business & Economic Statistics*, **29** (2), pp. 238–249.

Card, D. (1999). The Causal Effect of Education on Earnings, in O. Ashenfelter and D. Card (Eds.) *Handbook of Labor Economics*, **3**, Handbook of Labor Economics: Elsevier, pp. 1801–1863.

Carneiro, P., C. Crawford, and A. Goodman (2007). The Impact of Early Cognitive and Non-Cognitive Skills on Later Outcomes, *CEE Discussion Papers*.

Carrell, S. E., T. Maghakian, and J. E. West (2011). A's from Zzzz's? The Causal Effect of School Start Time on the Academic Achievement of Adolescents, *American Economic Journal: Economic Policy*, **3** (3), pp. 62–81.

Carroll, G. D., J. J. Choi, D. Laibson, B. C. Madrian, and A. Metrick (2009). Optimal Defaults and Active Decisions, *Quarterly Journal of Economics*, **124** (4), pp. 1639–1674.

Casey, B. J., L. H. Somerville, I. H. Gotlib, O. Ayduk, N. T. Franklin, M. K. Askren, J. Jonides, M. G. Berman, N. L. Wilson, T. Teslovich, G. Glover, V. Zayas, W. Mischel, and Y. Shoda (2011). Behavioral and neural correlates of delay of gratification 40 years later, *Proceedings of the National Academy of Sciences of the United States of America*, **108** (36), pp. 14998–15003.

Caskey, J. (2015). cgmwildboot.ado.

Castillo, M., P. J. Ferraro, J. L. Jordan, and R. Petrie (2011). The today and tomorrow of kids: Time preferences and educational outcomes of children, *Journal of Public Economics*, **95** (11-12), pp. 1377–1385.

Castillo, M., J. L. Jordan, and R. Petrie (forthcoming). Discount rates of children and high school graduation, *Economic Journal*.

Chetty, R., A. Looney, and K. Kroft (2009). Salience and taxation: Theory and evidence, *American Economic Review*, **99** (4), pp. 1145–1177.

Clark, D. (2009). The Performance and Competitive Effects of School Autonomy, *Journal of Political Economy*, **117** (4), pp. 745–783.

Clark, D. and H. Royer (2013). The Effect of Education on Adult Mortality and Health: Evidence from Britain, *American Economic Review*, **103** (6), pp. 2087–2120.

Compas, B. and B. Wagner (1991). Psychosocial stress during adolescence: Intrapersonal and interpersonal processes, in M. E. Colten S. Gore (Ed.) *Adolescent stress: Causes and consequences*, Hawthorne, NY, US: Aldine de Gruyter, pp. 67–85.

Costello, E. J., W. Copeland, and A. Angold (2011). Trends in psychopathology across the adolescent years: what changes when children become adolescents, and when adolescents become adults?, *Journal of child psychology and psychiatry, and allied disciplines*, **52** (10), pp. 1015–1025.

Courtemanche, C., G. Heutel, and P. McAlvanah (2015). Impatience, Incentives and Obesity, *Economic Journal*, **125** (582), pp. 1–31.

Cradock, A. L., S. J. Melly, J. G. Allen, J. S. Morris, and S. L. Gortmaker (2007). Characteristics of School Campuses and Physical Activity Among Youth, *American Journal of Preventive Medicine*, **33** (2), pp. 106 – 113.

Crespo, L., B. López-Noval, and P. Mira (2014). Compulsory schooling, education, depression and memory: New evidence from SHARELIFE, *Economics of Education Review*, **43**, pp. 36–46.

Crossley, T. F. and S. Kennedy (2002). The reliability of self-assessed health status, *Journal of Health Economics*, **21** (4), pp. 643–658.

Crystal, D., C. Chen, A. Fuligni, H. Stevenson, C.-C. Hsu, H.-J. Ko, S. Kitamura, and S. Kimura (1994). Psychological Maladjustment and Academic Achievement: A Cross-Cultural Study of Japanese, Chinese, and American High School Students, *Child Development*, **65** (3), pp. pp. 738–753.

Cunha, F. and J. Heckman (2007). The Technology of Skill Formation, *American Economic Review*, **97** (2), pp. 31–47.

Cunha, F. and J. J. Heckman (2008). Formulating, Identifying and Estimating the Technology of Cognitive and Noncognitive Skill Formation, *Journal of Human Resources*, **43** (4), pp. 738–782.

Currie, J. and E. Moretti (2003). Mother's education and the intergenerational transmission of human capital: Evidence from college openings, *Quarterly Journal of Economics*, **118** (4), pp. 1495–1532.

Cutler, D. M. and A. Lleras-Muney (2010). Understanding differences in health behaviors by education, *Journal of Health Economics*, **29** (1), pp. 1–28.

Dahmann, S. C. and S. Anger (2014). The Impact of Education on Personality: Evidence from a German High School Reform, *IZA Discussion Paper No. 8139*.

Dahmann, S. C. (in press). How Does Education Improve Cognitive Skills? Instructional Time versus Timing of Instruction, *Labour Economics*.

Dam, H. (2005). Der neue Lehrplan für das verkürzte Gymnasium, *Schönberger Hefte*, **35** (4), pp. 14–17.

De Neve, J.-E. and A. Oswald (2012). Estimating the influence of life satisfaction and positive affect on later income using sibling fixed effects, *Proceedings of the National Academy of Sciences*, **109** (49), pp. 19953–19958.

Dee, T. S. (2004). Teachers, Race, and Student Achievement in a Randomized Experiment, *Review of Economics and Statistics*, **86** (1), pp. 195–210.

Dohmen, T., A. Falk, D. Huffman, and U. Sunde (2010). Are Risk Aversion and Impatience Related to Cognitive Ability?, *American Economic Review*, **100** (3), pp. 1238–1260.

Dornbusch, S., R. Mont-Reynaud, P. Ritter, Z.-Y. Chen, and L. Steinberg (1991). *Stressful events and their correlates among adolescents of diverse backgrounds*, pp. 111–130, Social institutions and social change. Hawthorne, NY, US: Aldine de Gruyter.

Dörsam, M. and V. Lauber (2015). The Effect of a Compressed High School Curriculum on University Performance, *Beiträge zur Jahrestagung des Vereins für Socialpolitik 2015: Ökonomische Entwicklung - Theorie und Politik - Session: Education 2, No. B07-V3.*

Duckworth, A. L. and M. E. Seligman (2005). Self-discipline outdoes IQ in predicting academic performance of adolescents, *Psychological science*, **16** (12), pp. 939–944.

Duckworth, A. L. and D. S. Yeager (2015). Measurement Matters: Educational Researcher, *Educational Researcher*, **44** (4), pp. 237–251.

Dursun, B. and R. Cesur (2016). Transforming lives: the impact of compulsory schooling on hope and happiness, *Journal of Population Economics*, **29** (3), pp. 911–956.

Dzewaltowski, D. A., P. A. Estabrooks, G. Welk, J. Hill, G. Milliken, K. Karteroliotis, and J. A. Johnston (2009). Healthy Youth Places: A Randomized Controlled Trial to Determine the Effectiveness of Facilitating Adult and Youth Leaders to Promote Physical Activity and Fruit and Vegetable Consumption in Middle Schools, *Health Education & Behavior*, **36** (3), pp. 583–600.

Edwards, F. (2012). Early to Rise? The Effect of Daily Start Times on Academic Performance, *Economics of Education Review*, **31** (6), pp. 970–983.

Eide, E. and M. H. Showalter (1998). The effect of school quality on student performance: A quantile regression approach, *Economics Letters*, **58** (3), pp. 345 – 350.

Eigsti, I.-M., V. Zayas, W. Mischel, Y. Shoda, O. Ayduk, M. B. Dadlani, M. C. Davidson, J. Lawrence Aber, and B. J. Casey (2006). Predicting cognitive control from preschool to late adolescence and young adulthood, *Psychological science*, **17** (6), pp. 478–484.

Elias, M. J. (1989). Schools as a source of stress to children: An analysis of causal and ameliorative influences, *Journal of School Psychology*, **27** (4), pp. 393–407.

Etilé, F. and C. Milcent (2006). Income-related reporting heterogeneity in self-assessed health: evidence from France, *Health Economics*, **15** (9), pp. 965–981.

Falk, A. and F. Kosse (2016). Early Childhood Environment, Breastfeeding and the Formation of Preferences, *HCEO Working Paper*, **2016-036**.

Fischer, M., M. Karlsson, and T. Nilsson (2013). Effects of compulsory schooling on mortality: evidence from Sweden, *International Journal of Environmental Research and Public Health*, **10** (8), pp. 3596–3618.

Forbes, H., E. Fichera, A. Rogers, and M. Sutton (2017). The effects of exercise and relaxation on health and wellbeing, *Health Economics, in press.*

Forehand, R., B. Neighbors, and M. Wierson (1991). The Transition to Adolescence: the Role of Gender and Stress in Problem Behavior and Competence, *Journal of Child Psychology and Psychiatry*, **32** (6), pp. 929–937.

Frederick, S., G. Loewenstein, and T. O'Donoghue (2002). Time Discounting and Time Preference: A Critical Review, *Journal of Economic Literature*, **40** (2), pp. 351–401.

Frisvold, D. and E. Golberstein (2011). School quality and the education–health relationship: Evidence from Blacks in segregated schools, *Journal of Health Economics*, **30** (6), pp. 1232–1245.

Fuchs, V. (1982). Time Preference and Health: An Exploratory Study, in *Economic Aspects of Health*: National Bureau of Economic Research, Inc, pp. 93–120.

Garrison, C. Z., M. D. Schluchter, V. J. Schoenbach, and B. K. Kaplan (1989). Epidemiology of depressive symptoms in young adolescents, *Journal of the American Academy of Child and Adolescent Psychiatry*, **28** (3), pp. 343–351.

Gathmann, C., H. Jürges, and S. Reinhold (2015). Compulsory schooling reforms, education and mortality in twentieth century Europe, *Social Science & Medicine*, **127**, pp. 74–82.

Gill, S. C., P. Butterworth, B. Rodgers, and A. Mackinnon (2007). Validity of the mental health component scale of the 12-item Short-Form Health Survey (MCS-12) as measure of common mental disorders in the general population, *Psychiatry Research*, **152** (1), pp. 63–71.

Global BMI Mortality Collaboration (2016). Body-mass index and all-cause mortality: individual-participant-data meta-analysis of 239 prospective studies in four continents, *The Lancet*, **388** (10046), pp. 776–786.

Gneezy, U., M. Niederle, and A. Rustichini (2003). Performance in Competitive Environments: Gender Differences, *Quarterly Journal of Economics*, **118** (3), pp. 1049–1074.

Gneezy, U. and A. Rustichini (2004). Gender and Competition at a Young Age, *American Economic Review*, **94** (2), pp. 377–381.

Golsteyn, B. H., H. Grönqvist, and L. Lindahl (2014). Adolescent Time Preferences Predict Lifetime Outcomes, *Economic Journal*, **124** (580), pp. F739–F761.

Grant, K., B. Compas, A. Stuhlmacher, A. Thurm, S. McMahon, and J. Halpert (2003). Stressors and child and adolescent psychopathology: Moving from markers to mechanisms of risk, *Psychological Bulletin*, **129** (3), pp. 447–466.

Grossman, M. (1972). On the concept of health capital and the demand for health, *Journal of Political Economy*, **80** (2), pp. 223–255.

——— (2015). The Relationship between Health and Schooling: What's New? *Nordic Journal of Health Economics*, **3** (1), pp. 7–17.

Haberkorn, K. and S. Pohl (2013). *Cognitive Basic Skills – Data in the Scientific Use File*, Bamberg: University of Bamberg, National Educational Panel Study.

Hampson, S. E., L. R. Goldberg, T. M. Vogt, and J. P. Dubanoski (2007). Mechanisms by which childhood personality traits influence adult health status: Educational attainment and healthy behaviors, *Health psychology : official journal of the Division of Health Psychology, American Psychological Association*, **26** (1), pp. 121–125.

Hanushek, E. A. (1999). The Evidence on Class Size, in S. E. Mayer and P. E. Peterson (Eds.) *Earning and learning*, Washington and D.C and New York and N.Y: Brookings Institution Press and Russell Sage Foundation, pp. 131–168.

——— (2002). Publicly provided education, in *Handbook of Public Economics*, **4**: Elsevier, pp. 2045 – 2141.

——— (2006). School Resources, in E. A. Hanushek and F. Welch (Eds.) *Handbook of the Economics of Education*, **2**: Elsevier, pp. 865–908.

Hanushek, E. A. and S. G. Rivkin (2006). Teacher quality, in E. A. Hanushek and F. Welch (Eds.) *Handbook of the Economics of Education*, **2**: Elsevier, pp. 1051–1078.

——— (2010). Generalizations about using value-added measures of teacher quality, *American Economic Review*, **100** (2), pp. 267–71.

Hanushek, E. A. and L. Wößmann (2006). Does educational tracking affect performance and inequality? Differences-in-differences evidence across countries, *Economic Journal*, **116** (510), pp. C63–C76.

——— (2008). The Role of Cognitive Skills in Economic Development, *Journal of Economic Literature*, **46** (3), pp. 607–668.

——— (2012). Do better schools lead to more growth? Cognitive skills, economic outcomes, and causation, *Journal of Economic Growth*, **17** (4), pp. 267–321.

Heckman, J. J. (2006). Skill Formation and the Economics of Investing in Disadvantaged Children, *Science*, **312** (5782), pp. 1900–1902.

Heckman, J., R. Pinto, and P. Savelyev (2013). Understanding the Mechanisms through Which an Influential Early Childhood Program Boosted Adult Outcomes, *American Economic Review*, **103** (6), pp. 2052–86.

Heineck, G. and S. Anger (2010). The returns to cognitive abilities and personality traits in Germany, *Labour Economics*, **17** (3), pp. 535–546.

Heissel, J. and S. Norris (2017). Rise and shine: The effect of school start times on academic performance from childhood through puberty, *Journal of Human Resources*, pp. 0815–7346R1.

Heller, K. (2002). *Begabtenförderung im Gymnasium: Ergebnisse einer zehnjährigen Längsschnittstudie*, Opladen: Leske und Budrich.

Herzhoff, K., S. C. Kushner, and J. L. Tackett (2017). 2 - Personality development in childhood, in J. Specht (Ed.) *Personality Development Across the Lifespan*: Academic Press, pp. 9 – 23.

Hinrichs, P. (2011). When the Bell Tolls: The Effects of School Starting Times on Academic Achievement, *Education Finance and Policy*, **6** (4), pp. 486–507.

Hofmann, S. and A. Mühlenweg (2017). Learning Intensity Effects in Students' Mental and Physical Health - Evidence from a Large Scale Natural Experiment in Germany, *Hannover Economic Papers (HEP) No. 622*.

Homuth, C. (2017). *Die G8-Reform in Deutschland: Auswirkungen auf Schülerleistungen und Bildungsungleichheit*, Wiesbaden: Springer VS.

Hoyle, R. and E. Davisson (2016). Varieties of self-control and their personality correlates, in K. D. Vohs and R. F. Baumeister (Eds.) *Handbook of Self-Regulation: Research, Theory, and Applications*: Guilford Publications, pp. 396 – 413.

Hübner, N., W. Wagner, J. Kramer, B. Nagengast, and U. Trautwein (2017). Die G8-Reform in Baden-Württemberg: Kompetenzen, Wohlbefinden und Freizeitverhalten vor und nach der Reform, *Zeitschrift für Erziehungswissenschaft*, **1** (2), p. 1.

Huebener, M., S. Kuger, and J. Marcus (in press). Increased instruction hours and the widening gap in student performance, *Labour Economics*.

Huebener, M. and J. Marcus (2015). Moving up a Gear: The Impact of Compressing Instructional Time into Fewer Years of Schooling, *DIW Discussion Papers* (1450), URL: http://ideas.repec.org/p/diw/diwwpp/dp1450.html.

——— (2017). Compressing instruction time into fewer years of schooling and the impact on student performance, *Economics of Education Review*, **58**, pp. 1–14.

Hungerford, T. and G. Solon (1987). Sheepskin Effects in the Returns to Education, *Review of Economics and Statistics*, **69** (1), pp. 175–177.

Idler, E. L. and Y. Benyamini (1997). Self-rated health and mortality: a review of twenty-seven community studies, *Journal of Health and Social Behavior*, pp. 21–37.

Imbens, G. W. and J. M. Wooldridge (2009). Recent Developments in the Econometrics of Program Evaluation, *Journal of Economic Literature*, **47** (1), pp. 5–86.

Jaeger, D. A. and M. E. Page (1996). Degrees Matter: New Evidence on Sheepskin Effects in the Returns to Education, *Review of Economics and Statistics*, **78** (4), pp. 733–740.

Kaiser, A. and R. Kaiser (1998). *Entwicklung und Erprobung von Modellen der Begabtenförderung am Gymnasium mit Verkürzung der Schulzeit: Abschlussuntersuchung in der Gymnasialen Oberstufe (MSS)*, Schulversuche und Bildungsforschung, Mainz: V. Hase und Koehler.

Karakurt, Y. (2011). Mein Kopf ist voll! Selbst gute Schüler wollen lieber länger lernen, *Die Zeit*, **34**.

Kemptner, D., H. Jürges, and S. Reinhold (2011). Changes in compulsory schooling and the causal effect of education on health: Evidence from Germany, *Journal of Health Economics*, **30** (2), pp. 340–354.

Krashinsky, H. (2014). How Would One Extra Year of High School Affect Academic Performance in University? Evidence from an Educational Policy Change, *Canadian Journal of Economics/Revue canadienne d'économique*, **47** (1), pp. 70–97.

Kühn, S., I. Ackeren, G. Bellenberg, C. Reintjes, and G. Brahm (2013). Wie viele Schuljahre bis zum Abitur? *Zeitschrift für Erziehungswissenschaft*, **16** (1), pp. 115–136.

Kühn, S. M. (2014). Sind 12 Schuljahre ausreichend für den Zugang zur Hochschule? Der doppelte Abiturjahrgang aus empirischer Perspektive, *Beiträge zur Hochschulforschung*, **36** (3), pp. 8–33.

Kultusministerkonferenz (2013). *Vereinbarung zur Gestaltung der gymnasialen Oberstufe in der Sekundarstufe II*: Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland. Original July 7, 1972, revised June 6, 2013, URL: http://www.kmk.org/fileadmin/veroeffentlichungen_beschluesse/1972/1972_07_07-Vereinbarung-Gestaltung-Sek2.pdf.

Langenfeld, T. E., S. Milner, and P. W. Veljkov (1997). The Effects of Children's Ability to Delay Gratification on School Related Behaviors., *Paper presented at the Annual Meeting of the American Education Research Assosciation, Chicago.*

Lavy, V. (2015). Do Differences in Schools' Instruction Time Explain International Achievement Gaps? Evidence from Developed and Developing Countries, *Economic Journal*, **125** (588), pp. F397–F424.

Lee, J. and R. J. Barro (2001). Schooling Quality in a Cross–Section of Countries, *Economica*, **68** (272), pp. 465–488.

Leibniz Institute for Educational Trajectories (2015). Starting Cohort 2: Kindergarten (SC2) Wave 3 Information on the Competence Test.

Lemmon, K. and C. Moore (2007). The development of prudence in the face of varying future rewards, *Developmental Science*, **10** (4), pp. 502–511.

Lévy-Garboua, L., Y. Lohéac, and B. Fayolle (2006). Preference formation, school dissatisfaction and risky behavior of adolescents, *Journal of Economic Psychology*, **27** (1), pp. 165–183.

Lewinsohn, P. M., P. Rohde, and J. R. Seeley (1998). Major depressive disorder in older adolescents: Prevalence, risk factors, and clinical implications, *Clinical Psychology Review*, **18** (7), pp. 765–794.

Lleras-Muney, A. (2005). The relationship between education and adult mortality in the United States, *The Review of Economic Studies*, **72** (1), pp. 189–221.

Lochner, L. and E. Moretti (2004). The effect of education on crime: Evidence from prison inmates, arrests, and self-reports, *American Economic Review*, **94** (1), pp. 155–189.

Lohmar, B. and T. Eckhardt (2013). *The Education System in the Federal Republic of Germany 2011/2012: A Description of the Responsibilities, Structures and Developments in Education Policy for the Exchange of Information in Europe*: Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland.

Lorenz, C., M. Schmitt, N. Luplow, and C. Schönmoser (2016). Soziale Disparitäten im Vorschulalter und der Einfluss der Selbstregulation, *Zeitschrift für Grundschulforschung : ZfG ; Bildung im Elementar- und Primarbereich*, **9** (1), pp. 65–77.

Lundborg, P., A. Nilsson, and D.-O. Rooth (2014). Adolescent health and adult labor market outcomes, *Journal of Health Economics*, **37**, pp. 25 – 40.

Machin, S., O. Marie, and S. Vujić (2011). The crime reducing effect of education, *Economic Journal*, **121** (552), pp. 463–484.

Marcus, J. and V. Zambre (forthcoming). The Effect of Increasing Education Efficiency on University Enrollment: Evidence from Administrative Data and an Unusual Schooling Reform in Germany.

Mazumder, B. (2008). Does education improve health? A reexamination of the evidence from compulsory schooling laws, *Economic Perspectives*, **32** (2), pp. 2–17.

Mazzonna, F. (2014). The long lasting effects of education on old age health: evidence of gender differences, *Social Science & Medicine*, **101**, pp. 129–138.

McCrae, R. R. and O. P. John (1992). An Introduction to the Five-Factor Model and Its Applications, *Journal of Personality*, **60** (2), pp. 175–215.

Meghir, C., M. Palme, and E. Simeonova (2012). Education, health and mortality: Evidence from a social experiment, *NBER Working Paper No. 17932*.

Merikangas, K. R., J.-P. He, M. Burstein, S. A. Swanson, S. Avenevoli, L. Cui, C. Benjet, K. Georgiades, and J. Swendsen (2010). Lifetime prevalence of mental disorders in U.S. adolescents: results from the National Comorbidity Survey Replication–Adolescent Supplement (NCS-A), *Journal of the American Academy of Child and Adolescent Psychiatry*, **49** (10), pp. 980–989.

Merikangas, K. R., E. F. Nakamura, and R. C. Kessler (2009). Epidemiology of mental disorders in children and adolescents, *Dialogues in Clinical Neuroscience*, **11** (1), pp. 7–20.

Meyer, T. and S. L. Thomsen (2015). Schneller fertig, aber weniger Freizeit? Eine Evaluation der Wirkungen der verkürzten Gymnasialschulzeit auf die außerschulischen Aktivitäten der Schülerinnen und Schüler, *Schmollers Jahrbuch*, **135** (3), pp. 249–277.

——— (2016). How Important Is Secondary School Duration for Postsecondary Education Decisions? Evidence from a Natural Experiment, *Journal of Human Capital*, **10** (1), pp. 67–108.

——— (2017). The role of high-school duration for university students' motivation, abilities and achievements, *Education Economics*, **4** (2), pp. 1–22.

Meyer, T., S. L. Thomsen, and H. Schneider (forthcoming). New Evidence on the Effects of the Shortened School Duration in the German States: An Evaluation of Post-School Education Decisions, *German Economic Review*.

Milde-Busch, A., A. Blaschek, I. Borggräfe, R. v. Kries, A. Straube, and F. Heinen (2010). Besteht ein Zusammenhang zwischen der verkürzten Gymnasialzeit und Kopfschmerzen und gesundheitlichen Belastungen bei Schülern im Jugendalter? *Klinische Pädiatrie*, **222** (04), pp. 255–260.

Milligan, K., E. Moretti, and P. Oreopoulos (2004). Does education improve citizenship? Evidence from the United States and the United Kingdom, *Journal of Public Economics*, **88** (9-10), pp. 1667–1695.

Mincer, J. (1958). Investment in Human Capital and Personal Income Distribution, *Journal of Political Economy*, **66** (4), pp. 281–302.

——— (1974). *Schooling, Experience, and Earnings*: National Bureau of Economic Research, New York.

Ministerium für Bildung Kultur und Wissenschaft, Saarland (2000). *Das achtjährige Gymnasium im Saarland Kürzere Schulzeit - bessere Chancen*, URL: www.saarland.de/dokumente/thema_bildung/brosch_g8.PDF.

Mischel, W., Y. Shoda, and M. I. Rodriguez (1989). Delay of gratification in children, *Science*, **244** (4907), p. 933.

Mischel, W. and E. B. Ebbesen (1970). Attention in delay of gratification, *Journal of Personality and Social Psychology*, **16** (2), pp. 329–337.

Mischel, W., E. B. Ebbesen, and A. Raskoff Zeiss (1972). Cognitive and attentional mechanisms in delay of gratification, *Journal of Personality and Social Psychology*, **21** (2), pp. 204–218.

Mischel, W. and C. Gilligan (1964). Delay of gratification, motivation for the prohibited gratification, and responses to temptation., *The Journal of Abnormal and Social Psychology*, **69** (4), p. 411.

Mischel, W. and R. Metzner (1962). Preference for delayed reward as a function of age, intelligence, and length of delay interval., *The Journal of Abnormal and Social Psychology*, **64** (6), p. 425.

Mischel, W., Y. Shoda, and P. K. Peake (1988). The nature of adolescent competencies predicted by preschool delay of gratification, *Journal of Personality and Social Psychology*, **54** (4), pp. 687–696.

Moffitt, T. E., L. Arseneault, D. Belsky, N. Dickson, R. J. Hancox, H. Harrington, R. Houts, R. Poulton, B. W. Roberts, S. Ross, M. R. Sears, W. M. Thomson, and A. Caspi (2011). A gradient of childhood self-control predicts health, wealth, and public safety, *Proceedings of the National Academy of Sciences of the United States of America*, **108** (7), pp. 2693–2698.

Montez, J. K. and E. M. Friedman (2015). Educational attainment and adult health: under what conditions is the association causal? *Social Science & Medicine* (127), pp. 1–7.

Morin, L.-P. (2015). Do Men and Women Respond Differently to Competition? Evidence from a Major Education Reform, *Journal of Labor Economics*, **33** (2), pp. 443–491.

Müller, D., T. Linberg, M. Bayer, T. Schneider, and F. Wohlkinger (2016). *Measuring Personality Traits of Young Children— Results From a NEPS Pilot Study*, pp. 169–180, Wiesbaden: Springer Fachmedien Wiesbaden, URL: https://doi.org/10.1007/978-3-658-11994-2_10.

Neubauer, A., C. Gawrilow, M. Hasselhorn, and W. Schneider (2011). Belohnungsaufschub: Ein Ansatz zur Frühprognose volitionaler Kompetenzen, *Frühprognose schulischer Kompetenzen*, pp. 203–220.

Neumann, I., C. Duchhardt, M. Grüßing, A. Heinze, E. Knopp, and T. Ehmke (2013). Modeling and assessing mathematical competence over the lifespan, *Journal for Educational Research Online / Journal für Bildungsforschung Online*, **5** (2).

Nolen-Hoeksema, S. and J. Girgus (1994). The emergence of gender differences in depression during adolescence, *Psychological bulletin*, **115** (3), p. 424—443.

Non, A. and D. Tempelaar (2016). Time preferences, study effort, and academic performance, *Economics of Education Review*, **54**, pp. 36–61.

OECD (2000). *Education at a Glance 2000*: OECD Publishing.

——— (2013a). *Education at a Glance 2013*: OECD Publishing.

——— (2013b). *PISA 2012 Results: Excellence through Equity (Volume II)*: OECD Publishing.

Oreopoulos, P. (2006). Estimating average and local average treatment effects of education when compulsory schooling laws really matter, *American Economic Review*, **96** (1), pp. 152–175.

Patel, V., A. J. Flisher, S. Hetrick, and P. McGorry (2007). Mental health of young people: A global public-health challenge, *The Lancet*, **369** (9569), pp. 1302–1313.

Patton, G. C., C. Coffey, H. Romaniuk, A. Mackinnon, J. B. Carlin, L. Degenhardt, C. A. Olsson, and P. Moran (2014). The prognosis of common mental disorders in adolescents: a 14-year prospective cohort study, *The Lancet*, **383** (9926), pp. 1404–1411.

Perez-Arce, F. (2017). The effect of education on time preferences, *Economics of Education Review*, **56**, pp. 52–64.

van Praag, B. and A. Ferrer-i Carbonell (2010). Happiness Economics: A New Road to Measuring and Comparing Happiness, *Foundations and Trends® in Microeconomics*, **6** (1), pp. 1–97.

Primi, R., M. E. Ferrão, and L. S. Almeida (2010). Fluid intelligence as a predictor of learning: A longitudinal multilevel approach applied to math, *Learning and Individual Differences*, **20** (5), pp. 446 – 451.

Quis, J. S. (2015). Does higher learning intensity affect student well-being? Evidence from the National Educational Panel Study, *BERG Working Paper Series* (94).

——— (fortcoming). Does Compressing High School Duration Affect Students' Stress and Mental Health? Evidence from the National Educational Panel Study, *Jahrbücher für Nationalökonomie und Statistik*.

Quis, J. S. and S. Reif (2017). Health effects of instruction intensity: Evidence from a natural experiment in German high-schools, *SOEP Papers* (916).

Raufelder, D., F. Kittler, S. R. Braun, A. Lätsch, R. P. Wilkinson, and F. Hoferichter (2013). The interplay of perceived stress, self-determination and school engagement in adolescence, *School Psychology International*, **35** (4), pp. 405–420.

Reilly, J. J. and J. Kelly (2011). Long-term impact of overweight and obesity in childhood and adolescence on morbidity and premature mortality in adulthood: systematic review, *International Journal of Obesity*, **35** (7), pp. 891–898.

Reivich, K., J. Gillham, T. Chaplin, and M. Seligman (2013). From Helplessness to Optimism: The Role of Resilience in Treating and Preventing Depression in Youth, in S. Goldstein and R. B. Brooks (Eds.) *Handbook of Resilience in Children*: Springer US, pp. 201–214.

Rivkin, S. G., E. A. Hanushek, and J. F. Kain (2005). Teachers, Schools, and Academic Achievement, *Econometrica*, **73** (2), pp. 417–458.

Rushton, J. L., M. Forcier, and R. M. Schectman (2002). Epidemiology of depressive symptoms in the National Longitudinal Study of Adolescent Health, *Journal of the American Academy of Child and Adolescent Psychiatry*, **41** (2), pp. 199–205.

Sallis, J. F., T. L. McKenzie, T. L. Conway, J. P. Elder, J. J. Prochaska, M. Brown, M. M. Zive, S. J. Marshall, and J. E. Alcaraz (2003). Environmental interventions for eating and physical activity: a randomized controlled trial in middle schools, *American Journal of Preventive Medicine*, **24** (3), pp. 209–217.

Schavan, A. and D. Ahnen (2001). Abitur nach 12 Schuljahren? Pro und Contra, *Forschung und Lehre*, **8** (9), pp. 472–473.

Schlam, T. R., N. L. Wilson, Y. Shoda, W. Mischel, and O. Ayduk (2013). Preschoolers' Delay of Gratification Predicts their Body Mass 30 Years Later, *Journal of Pediatrics*, **162** (1), pp. 90–93.

Schnittjer, I. (2018). NEPS Technical Report for Mathematics: Scaling Results of Starting Cohort 2 for Grade 1, *NEPS Survey Paper* (44).

Schnittjer, I. and C. Duchhardt (2015). *Mathematical Competence: Framework and Exemplary Test Items*, Bamberg: University of Bamberg, National Educational Panel Study.

Schnittjer, I. and A.-L. Gerken (2018). NEPS Technical Report for Mathematics: Scaling Results of Starting Cohort 2 for Grade 2, *NEPS Survey Paper* (47).

Schönberger, B. and C. Aßmann (2014). Weighting the Additional Study in Baden-Wuerttemberg of the National Educational Panel Study, URL: https://www.neps-data.de/Portals/0/NEPS/Datenzentrum/Forschungsdaten/BW/2-0-0/BW_2-0-0_Weighting.pdf.

Schultz, T. W. (1959). Investment in Man: An Economist's View, *Social Service Review*, **33** (2), pp. 109–117.

——— (1960). Capital formation by education, *Journal of Political Economy*, **68** (6), pp. 571–583.

——— (1961). Investment in Human Capital, *American Economic Review*, **51** (1), pp. 1–17.

——— (1962). Reflections on investment in man, *Journal of Political Economy*, **70** (5, Part 2), pp. 1–8.

Schutte, N. S., J. M. Malouff, E. Segrera, A. Wolf, and L. Rodgers (2003). States reflecting the Big Five dimensions, *Personality and Individual Differences*, **34** (4), pp. 591–603.

Scrimin, S., U. Moscardino, G. Altoè, and L. Mason (2016). Effects of perceived school well-being and negative emotionality on students' attentional bias for academic stressors, *British Journal of Educational Psychology*, **86** (2), pp. 278–295.

Seeyave, D. M., S. Coleman, D. Appugliese, R. F. Corwyn, R. H. Bradley, N. S. Davidson, N. Kaciroti, and J. C. Lumeng (2009). Ability to delay gratification at age 4 years and risk of overweight at age 11 years, *Archives of Pediatrics & Adolescent Medicine*, **163** (4), pp. 303–308.

Seiffge-Krenke, I. (2000). Causal links between stressful events, coping style, and adolescent symptomatology, *Journal of Adolescence*, **23** (6), pp. 675 – 691.

——— (2007). Depression bei Kindern und Jugendlichen: Prävalenz, Diagnostik, ätiologische Faktoren, Geschlechtsunterschiede, therapeutische Ansätze, *Praxis der Kinderpsychologie und Kinderpsychiatrie*, **56**, pp. 185–205.

Sethi, A., W. Mischel, J. L. Aber, Y. Shoda, and M. L. Rodriguez (2000). The role of strategic attention deployment in development of self-regulation: Predicting preschoolers' delay of gratification from mother–toddler interactions., *Developmental Psychology*, **36** (6), p. 767.

Shamosh, N. A. and J. R. Gray (2008). Delay discounting and intelligence: A meta-analysis, *Intelligence*, **36** (4), pp. 289–305.

Shoda, Y., W. Mischel, and P. K. Peake (1990). Predicting adolescent cognitive and self-regulatory competencies from preschool delay of gratification: Identifying diagnostic conditions, *Developmental Psychology*, **26** (6), pp. 978–986.

Siedler, T. (2010). Schooling and citizenship in a young democracy: evidence from postwar Germany, *Scandinavian Journal of Economics*, **112** (2), pp. 315–338.

Silles, M. A. (2009). The causal effect of education on health: Evidence from the United Kingdom, *Economics of Education Review*, **28** (1), pp. 122–128.

Skirbekk, V. (2006). Does School Duration Affect Student Performance? Findings from Canton-based Variation in Swiss Educational Length, *Schweizerische Zeitschrift für Volkswirtschaft und Statistik*, **142** (I), pp. 115–145.

Smith, J. P. and G. C. Smith (2010). Long-term economic costs of psychological problems during childhood, *Social Science & Medicine*, **71** (1), pp. 110–115.

SOEP (2015). Socio-Economic Panel, data for years 1984-2014, *version 31, doi:10.5684/soep.v31.*

Spence, M. (1973). Job Market Signaling, *Quarterly Journal of Economics*, **87** (3), pp. 355–374.

Statistisches Bundesamt (2013). Fachserie 11/Reihe 1, Bildung und Kultur. Allgemeinbildende Schulen., Wiesbaden, URL: www.destatis.de/DE/Publikationen/Thematisch/BildungForschungKultur/Schulen/AllgemeinbildendeSchulen2110100137004.pdf?__blob=publicationFile.

——— (2014). Fachserie 11/Reihe 1, Bildung und Kultur. Allgemeinbildende Schulen., Wiesbaden, URL: https://www.destatis.de/GPStatistik/servlets/MCRFileNodeServlet/DEHeft_derivate_00013704/2110100147004_10112014.pdf.

Sussebach, H. (2011). Liebe Marie, *Die Zeit*, **22**, pp. 15–18.

Sutter, M., L. Yilmaz, and M. Oberauer (2015). Delay of gratification and the role of defaults—An experiment with kindergarten children, *Economics Letters*, **137**, pp. 21–24.

Thapar, A., S. Collishaw, D. S. Pine, and A. K. Thapar (2012). Depression in adolescence, *The Lancet*, **379** (9820), pp. 1056–1067.

Thiel, H., S. L. Thomsen, and B. Büttner (2014). Variation of learning intensity in late adolescence and the effect on personality traits, *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, **177** (4), pp. 861–892.

Thomas, R. E., J. McLellan, and R. Perera (2013). School-based programmes for preventing smoking, *Evidence-Based Child Health: A Cochrane Review Journal*, **8** (5), pp. 1616–2040.

Thomsen, S. (2015). The impacts of shortening secondary school duration, *IZA World of Labor*, p. 166.

Torres, S. J. and C. A. Nowson (2007). Relationship between stress, eating behavior, and obesity, *Nutrition*, **23** (11), pp. 887–894.

Van Doorslaer, E. and U.-G. Gerdtham (2003). Does inequality in self-assessed health predict inequality in survival by income? Evidence from Swedish data, *Social Science & Medicine*, **57** (9), pp. 1621–1629.

Van Kippersluis, H., O. O'Donnell, and E. Van Doorslaer (2011). Long-run returns to education does schooling lead to an extended old age? *Journal of Human Resources*, **46** (4), pp. 695–721.

Vohs, K. D. and R. F. Baumeister (2016). *Handbook of self-regulation: Research, theory, and applications*: Guilford Publications, 3rd edition.

Wagner, G. G., J. R. Frick, and J. Schupp (2007). The German Socio-Economic Panel study (SOEP)-evolution, scope and enhancements, *Schmollers Jahrbuch*, **127** (1), pp. 139–170.

Wagner, W., J. Kramer, U. Trautwein, O. Lüdtke, G. Nagy, K. Jonkmann, K. Maaz, S. Meixner, and J. Schilling (2011). Upper secondary education in academic school tracks and the transition from school to postsecondary education and the job market, *Zeitschrift für Erziehungswissenschaft*, **14** (2), pp. 233–249.

de Walque, D. (2007). Does education affect smoking behaviors?: Evidence using the Vietnam draft as an instrument for college education, *Journal of Health Economics*, **26** (5), pp. 877 – 895.

——— (2010). Education, information, and smoking decisions evidence from smoking histories in the United States, 1940–2000, *Journal of Human Resources*, **45** (3), pp. 682–717.

Watts, T., G. Duncan, and H. Quan (2018). Revisiting the Marshmallow Test: A Conceptual Replication Investigating Links Between Early Gratification Delay and Later Outcomes, *Psychological Science*, **29**.

Watts, T. W., G. J. Duncan, D. H. Clements, and J. Sarama (2017). What Is the Long-Run Impact of Learning Mathematics During Preschool? *Child Development*, pp. 539–555.

Watts, T. W., G. J. Duncan, R. S. Siegler, and P. E. Davis-Kean (2014). What's Past Is Prologue: Relations Between Early Mathematics Knowledge and High School Achievement, *Educational Researcher*, **43** (7), pp. 352–360.

Westermaier, F. (2016). The Impact of Lengthening the School Day on Substance Abuse and Crime: Evidence from a German High School Reform, *DIW Discussion Paper No. 1616*.

Wiater, W. (1996). Zwölf Jahre bis zum Abitur? Positionen im Streit um die Verkürzung der gymnasialen Schulzeit, in W. Melzer and U. Sandfuchs (Eds.) *Schulreform in der Mitte der 90er Jahre*, **8**, Reihe Schule und Gesellschaft: VS Verlag für Sozialwissenschaften, pp. 121–139.

——— (1997). Abitur nach 12 oder 13 Jahren? Die Diskussion um die Schulzeitverkürzung und ihre Folgen, *Praxis Schule*, **8** (3), pp. 5–10.

Zydatiß, W. (1999). Förderung über Akzeleration: Gymnasiale Express-und Regelklassen im Vergleich, *Schulverwaltung MO*, **7**, pp. 255–260.