Learning Intensity Effects in Students' Mental and Physical Health – Evidence from a Large Scale Natural Experiment in Germany

Sarah Hofmann[*] & Andrea Mühlenweg[†]

December 6, 2017

Abstract

In this study, we analyze health effects of a recent education reform in Germany exposing students to increased schooling intensity. The reform shortened the higher secondary education track by one year. As the overall curriculum required for graduation was held constant, this led to an increase in instruction hours in the remaining school years. The reform has been introduced at different points in time across federal states, providing us with a difference-in-difference setup for analysis. Based on data from the German Socio-Economic Panel Study (SOEP), our results imply that the reform significantly reduced adolescents' self-rated mental health status. The overall effect on the mental component summary score (MCS) is about a quarter of a standard deviation. Examining MCS sub-dimensions, we find detrimental effects of the reform on vitality and on emotional balance. We also observe significant impacts on self-assessed general physical health.

JEL Classification: J24, I14, I26

Keywords: Adolescent health, schooling intensity, school reform, natural experiment

1 Introduction

Mental health problems and detrimental health behaviors among adolescents are on the rise in many countries, thus generating major public health policy concerns (WHO, 2017; Mannan et al., 2016). According to recent data published by the World Health Organization, mental health problems are the predominant causes of adolescent disability-adjusted life years lost in high income countries (WHO, 2017). From a theoretical point of view, health in childhood and adolescence is key in the skill formation process that produces adult human capital (Heckman, 2007). Referring to the literature on skill formation, our paper adds empirical insight on the malleability of (mental) health in the school context (Cunha and Heckman, 2007; Heckman, 2007). We are able to identify a clear causal pathway from the school environment to health in adolescence. Specifically, we analyze a recent large scale reform in Germany which increased learning intensity in higher secondary schools. Observing affected and non-affected students ages 18-19 allows us to quantify learning intensity impacts on self-assessed (mental) health and health behavior at this age. Findings from our study shed light on the role of education policy at the secondary school level in causally shaping the development of health and health inequalities. More broadly, our paper contributes to the literature disentangling the interrelations of education and health (Cutler and Lleras-Muney, 2012; Feinstein et al., 2006; Grossman, 2015).

The G8 reform, implemented between 2001 and 2007, reduced the years of education required for the higher secondary degree from 13 to 12 years in Germany, i.e. from nine to eight years of secondary schooling after the four year elementary school. The reform affected students in academic track schools, which constitute the major secondary school track in Germany (*Gymnasium*).[1] As the overall curriculum required for the academic secondary degree was held constant, the reform led to an increased schooling intensity in the remaining school years (Huebener and Marcus, 2015a; Homuth, 2017). In the public debate, the latter feature is widely discussed as a reason for increased stress and negative mental health outcomes among young adults (Homuth, 2017). Accordingly, we suspect that higher learning intensity increases distress in secondary school students.[2]

To date, there is limited evidence on the impact of learning intensity on health. Pischke (2007) examines a historic large scale reform that aimed at a harmonization of school year starting dates in Germany. Similar to the reform considered in our paper, this historic reform decreased the time spent in school while the curriculum remained unchanged. Students in school in 1966-1967 were exposed to short school years. These students received about two thirds of a year less schooling compared to the previous years. Pischke (2007) makes use of schooling variation across cohorts, states and school tracks in order to identify

---

[1]Section 2 details features of the German school system and the reform. Students in Germany are generally streamed to three different types of secondary school tracks after elementary school. Compared to the *Gymnasium*, the *Realschule* and *Hauptschule* tracks provide more vocationally-oriented secondary education.

[2]School-related stress is one potential channel affecting students' mental health and might even affect physical health. Based on our data, we will also consider students' self-assessment of being under pressure as well as information on their leisure time use and social interaction.

reform-based effects. The study finds no adverse effects of compressed schooling time on labor market outcomes such as employment and earnings. However, students affected by the reform were more likely to repeat a grade. According to Pischke (2007), this finding points to detrimental learning effects at the bottom of the performance distribution. Concerning health effects, Braakmann (2010) reconsiders the same quasi-experimental setting as Pischke (2007) and draws on aggregated health indicators available in the German Socio-Economic Panel (SOEP). Based on this data, the study does not find robust long-run effects on weight or on self-assessed overall physical and mental health. Similar to Pischke (2007), Fischer et al. (2016) analyze large scale historic school reforms *increasing* compulsory schooling as well as term length in Sweden in the 1930s and 1940s. The study's difference-in-difference analyses point to strong earnings effects associated with the reforms. These beneficial effects are especially pronounced for females and at lower quantiles when term length is increased.

More recently, reforms similar to the German G8 reform were introduced in Switzerland in 1995 as well as in Ontario, Canada, in the period between 1999 and 2003. Evaluation of the Swiss reform is hampered by the fact that other (curricular) changes came into effect at the same time (Ramseier et al., 2004). The Ontario reform reduced university preparatory education from 13 years to 12 years while compressing the overall curriculum and holding the required amount of credits constant (cf. Thomsen, 2015). In a quasi-experimental study addressing this reform, Krashinsky (2014) draws on students from the double cohort, graduating in the same year under the old and new regime. Based on survey data on grades and academic performance, Krashinsky (2014) finds negative effects on performance at university. To our knowledge, there is no evidence on health effects of the Ontario reform.

Differences in the above findings from different reforms may be due to the populations affected: Historic reforms of compulsory schooling mainly affected students at a lower educational level (i.e. in compulsory schooling) while the Ontario reform as well as the G8 reform affect higher levels of secondary education.

Asides from the mentioned studies, further literature on school quality and school resources considers changes in instruction time without curricular changes as a measure for education quality. Patall et al. (2010) provides a meta-analysis on evidence related to variations in allocated school time due to extending the school day or school year. Designs are criticized as suffering from weak identification strategies and being mostly limited to performance effects. Still, the overall evidence indicates that more disadvantaged groups of students benefit from an increase in allocated school time.[3] However, a mere quality interpretation of higher learning intensity is not applicable to our case. In the context of the G8 reform, we suspect that there are counteracting effects of lower overall education years,

---

[3]In their seminal study on school resources, Card and Krueger (1992) do not consistently find effects of an increase in term length when considering monetary returns to education. In contrast, there are significant and robust effects for the alternative quality measure *pupil/teacher ratio*. More recently, for example Lavy (2016) examines a school funding reform in Israel and finds positive effects of increased weekly instruction hours on student test scores as well as on homework time. However, there were found no effects on school satisfaction or on school violence. Furthermore, a stream of this literature discusses increases in term length as feasible means in order to increase education outcomes in developing countries (Agüero and Beleche, 2013).

higher learning intensity and distress due to the corresponding workload. The latter effects will apply to students observed while being in school.

In the economic human capital literature, the *technology of capability (or skill) formation*, provides a theoretical framework on how different capabilities are shaped over the life-cycle. Capability production is modelled to be a multi-stage process implying critical and sensitive periods for the development of specific capabilities (Cunha and Heckman, 2007; Heckman, 2007). Skills from earlier periods are self-productive and complementary in the production of later period skills. This view extends the traditional Grossman model which considers adult health investment decisions without taking the life-cycle development into account (Grossman, 1972; Heckman, 2007). Empirical studies relating to the technology of skill formation have demonstrated that cognitive skills are relatively stable in adolescence, whereas non-cognitive dimensions of human capital are more malleable in the adolescence period (Borghans et al., 2008; Almlund et al., 2011). Thus, we also assume that the reform varying school inputs at the secondary education level may yield persistent (mental) health effects. Since cognitive and non-cognitive abilities as well as mental and physical health acquired in earlier periods positively affect skill formation at later stages, we also expect differential effects for students at differential positions of the skill distributions. This is in line with psychologic theory considering multiple risk factors that interact with stressors as well as protective factors in coping with distress over the life-cycle, e.g. according to the vulnerability-stress-model (Ingram and Luxton, 2005; Belsky and Pluess, 2009). More generally, in the psychological literature, school stress has been shown to be correlated to (mental) health outcomes as well as to health behaviors (e.g. Shankar and Park, 2016; Weidner et al., 1996). Particularly, regarding mental health, stress around schooling and achievement is related to conditions such as depressive episodes or affective disorders (Jacobshagen and Rigotti, 2008).

A range of recent studies analyze the effects of the G8 reform on personality traits (mainly looking at the big five inventory): Anger and Dahmann (2014) present evidence on the G8 reform affecting students' development of personality, based on the same difference-in-difference identification set-up as in our study. Their analysis of the SOEP data finds increases in extraversion as well as neuroticism in students under the reform (Anger and Dahmann, 2014). Thiel et al. (2014) study the effect of the reform on personality trait measures including locus of control and self-control. Based on survey data from twelve schools in the German state of Saxony-Anhalt and comparing double cohort graduates affected and not affected by the reform, they do not find effects on the considered skills about two years after graduation.[4] None of these studies directly addresses students' (mental) health outcomes.

There are some (correlative) studies, based on rather small samples, discussing how the reform might impact on health: Based on self-collected survey data from one German city used in a correlative analysis, Milde-Busch et al. (2010) find no systematic patterns in G8

---

[4]Similarly, based on a follow-up survey of the same data, Meyer and Thomsen (2016) find no overall differences in motivation for G8 and G9 in older students in the tertiary education context.

students' subjective health indicators such as stress, prevalence of pain and overall well-being. Further evidence on stress effects observed in a small sample of 69 students is presented in Minkley et al. (2015). This study shows that measured cortisol levels tend to be higher for G8 as compared to G9 double cohort graduates in the German state of North Rhine-Westphalia. Quis (2015) draws on data from only one German state to examine effects of the reform on an aggregated stress measure, on aggregated mental health, and on overall well-being. The study uses survey data of the German National Educational Panel Study (NEPS) and finds an increase in stress levels as well as a deterioration of mental health for girls under the reform. There are no effects on self-assessed well-being. The empirical framework including data on only one federal state (Baden-Württemberg) limits the causal interpretation of the results. Westermaier (2016) considers changes in illegal substance abuse interacting with the introduction of the G8 reform. Based on the federal states' crime reports, the study finds that the reform reduced illegal cannabis consumption. This might be interpreted as a specific measure of health behavior.[5]

When starting our work on this topic in 2016, we were not aware of any causal study using the reform's difference-in-difference setup in order to directly analyze health outcomes. However, at the time when we conducted our empirical research, coincidentally another team of authors studied the very same topic: Quis and Reif (2017) have recently come up with a discussion paper analyzing the effects of the G8 reform on some (but not all) of the outcomes which we consider in our paper. Given that – for the longest time of our work – we were not aware of their project being close to our own endeavor, we consider it legitimate to publish our paper as an original contribution to the literature that existed at that time. Our study results from longer discussions within our research network. While our research has been conducted independent from the Quis and Reif (2017) paper, in the following paragraph, we summarize their findings and briefly discuss how their work differs from our study.

Based on somewhat different samples from the SOEP than the ones we are using, Quis and Reif (2017) apply a triple difference strategy to assess the reform's impact on overall physical health, mental well-being and body mass index (BMI, not considered in our paper) of students and recent graduates.[6] Their "third difference" refers to comparing students in academic track schools and students in vocational track schools, where the latter are not directly affected by the reform. However, as we point out in Section 3 (empirical strategy), we consider our difference-in-difference strategy to be the more appropriate design. The triple difference approach is problematic as there might be simultaneous (indirect) effects of the reform on students in vocational track schools. This may also explain some of the insignificant findings in the study by Quis and Reif (2017), besides the explanation of smaller samples. Quis and Reif (2017) still find a negative effect of the reform on overall mental well-being and on BMI – a higher BMI driven by higher numbers of overweight persons –

---

[5]Further studies examine reform effects on student performance. We summarize this evidence in Section 2 together with more detail on the reform.

[6]Different samples are also due to the somewhat different selection of outcome variables, where observations with extreme BMI outcomes are dropped from the sample.

for females, however not for males, while they are in school. For graduates they do not find effects of the reform on the considered (aggregated) health outcomes.

In light of the previous evidence (at the time when we conducted our empirical work), our results add insight on the impact of learning intensity on detailed mental and physical health outcomes; we consider scores from the Physical and Mental Component Health Summary Scales (MCS, PCS) as well as all of their sub-scales. We also use indictors on health behavior (smoking and heavy smoking). We find significant reform effects on mental health which are robust to alternative specifications and samples. We further observe significant effects on the measure of self rated overall physical health. Regarding our summary score for physical health as well as the indicators on health behavior, our results are mostly insignificant. In sum, the identified effects are plausible, as we expect school stress to predominantly affect mental health and subjective health status (Jacobshagen and Rigotti (2008) on school stress). Our data also provide some information to address how the reform specifically affected students' daily life. According to this, the reform resulted in an increase in private tutoring as well as an increase in conflicts between parents and adolescents. While sub-group analyses are hampered by limited numbers of observations, our findings point to more pronounced effects for more vulnerable sub-groups. This is in line with the psychological theory of the diatheses-stress-model (Ingram and Luxton, 2005).

Our paper proceeds as follows: Section 2 details the education reform while Section 3 depicts how we use the timing of the reform in different states in our empirical identification strategy. Section 4 provides information on our database and outcome measures. Section 5 presents our results together with sub-group analyses and observed changes in students' daily activities. In Section 6, we further challenge our identification strategy and provide a series of robustness checks including placebo regressions. Section 7 discusses our findings and concludes.

2 Institutional Background

In Germany, education policy at the secondary schooling level is upon the federal states' responsibilities. According to the federal states' school laws, students are generally streamed to secondary track schools based on their performance in elementary school. This track choice is typically made when children are about ten years old, in the end of the fourth year in elementary school.[7] Higher performing students are supposed to be streamed to academic secondary schools (*Gymnasium*). In contrast, the intermediate track (*Realschule*) and short track (*Hauptschule*) are more oriented towards practical training. These tracks

---

[7]The four years in elementary school may be followed by an orientation stage for track choice up to grade six. Also, while tracking is principally supposed to be based on students' abilities, there may be parental and teacher discretion in track choices.

typically comprise grades 5-10 and 5-9, respectively. The curriculum of the academic track traditionally comprises nine years of schooling (grades 5-13).[8]

Among the traditional secondary school types, academic track schools directly prepare for the university qualification examination (*Abitur*). The *Haupt-* and *Realschule* tracks are typically followed by about three years of vocational training. Figure A-1 in the Appendix provides a stylized illustration of the German school tracks. As noted in Figure A-1, about 38 % of students are streamed to the academic track. Comprehensive or combined secondary track schools offer alternatives to the traditional segregated tracks but are not available all over Germany.

States did decide on the G8 reform without affected students having much of an option to avoid the G8 system (see Section 3 for further details and checks on deviating behavior): Decided upon in the beginning of the new century and implemented in most federal states in Germany between 2001 and 2007, the reform reduces the academic track school years from nine years (G9) to eight years (G8).

The reform is intended to more efficiently use the time that students spend in school and to reduce students' entry age to the labor market (Autorengruppe Bildungsberichterstattung, 2008). Under the G9 system, students and labor market entrants were documented to be relatively old in international comparison.[9]

While the reform reduced academic secondary schooling by one year, the overall curricular content was not reduced. This implies that schooling hours of the skipped year needed to be shifted to increase weekly schooling hours in the remaining years. Such a redistribution was possible, because in the G9 system, instruction hours were mostly in the morning (8 a.m. to 1 p.m.). Students typically spent the afternoon doing homework, attending tutoring classes, participating in sports or cultural clubs and doing other leisure time activities (Engels and Thielebein, 2011; Hille et al., 2016). According to Homuth (2017), average weekly instruction hours increased by three hours per school year (amounting to about 7% of the entire time spent at the academic track schools). More precisely, on average students had to attend two additional weekly hours in grades five as well as in grade six and about three more hours in each of the grades between grade seven to twelve (also based on figures in Homuth, 2017).[10]

---

[8]The school tracks allow for some degree of flexibility. In the academic track school, a lower secondary degree is obtained after successful completion of grade ten, in some states combined with specific examinations. Thus, higher track students may still leave the school track after grade ten and enter vocational training. On the other hand, vocational track students may later opt for (vocational) upper secondary school types and obtain a higher secondary degree to access Universities of Applied Sciences. However, switching tracks in the lower secondary education stage (up to grade 10) is relatively rare: According to Bellenberg (2012), depending on the state, between 1 % (Baden-Württemberg) and 6 % (Bremen, not in our sample) of students switch schools up to grade 10. Among all track changers, about 58 % switch to a lower type track.

[9]According to OECD (2002), at that time the median age of entrants to the tertiary education system was 21 years, which is higher than the median age in about two thirds of OECD countries with reported age distributions (total of 22 countries). Median entry age was lowest in Belgium (fl.), France, Ireland, Spain, the U.S. and U.K. (about 19 years).

[10]The increase in these minimum required hours is lower than an increase corresponding to an entire schooling year. An analysis of the different curricula points out that the corresponding reduction of instruction hours mostly

Figure 1: Implementation of G8 across federal states and graduation year of affected cohorts

Federal state / Year	2006	2007	2008	2009	2010	2011	2012	2013	2014
Saxony-Anhalt	G9	G8	G8	G8	G8	G8	G8	G8	G8
Mecklenburg West. Pom.	G9	G9	G8	G8	G8	G8	G8	G8	G8
Hamburg	G9	G9	G9	G9	G8	G8	G8	G8	G8
Bavaria	G9	G9	G9	G9	G9	G8	G8	G8	G8
Lower Saxony	G9	G9	G9	G9	G9	G8	G8	G8	G8
Baden-Württemberg	G9	G9	G9	G9	G9	G9	G8	G8	G8
Berlin	G9	G9	G9	G9	G9	G9	G8	G8	G8
North Rhine-Westphalia	G9	G9	G9	G9	G9	G9	G9	G8	G8

G9   G8
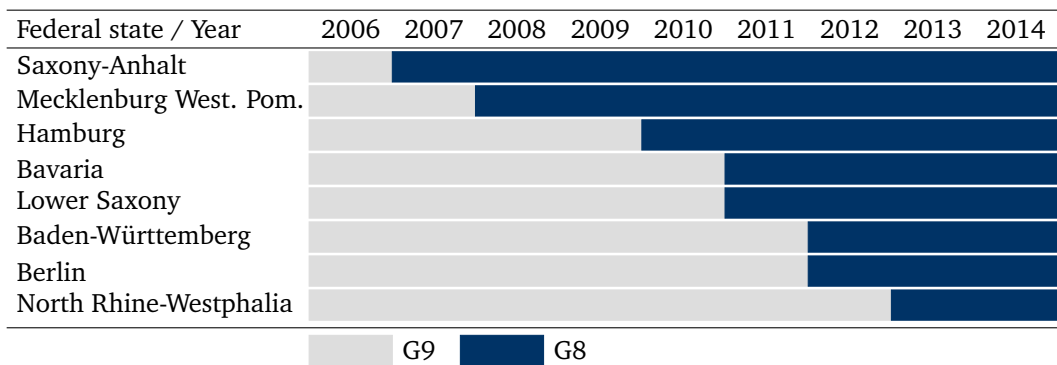
Source: Huebener and Marcus (2015a). Note: We only include states where we observe implementation of the reform in our sample.

The reform has been intensely debated among parents, teachers and policy-makers since its introduction. To date, most parents still seem to oppose G8 (cf. Wößmann et al., 2014). The main reasons for objecting to the reform are expected quality losses of secondary education and higher expected stress levels for students due to the rigid budgeting of time (Homuth, 2017).

There are several channels through which the reform may affect students' situation. For students in the first cohort affected, the reform implies that they graduate at the same time as the last G9 cohort (double cohorts). This results in tougher competition for later tertiary education and job applicants. To examine this implication, Meyer and Thomsen (2016) observe data for the double cohort graduates from 12 schools in the state of Saxony-Anhalt up to seven years after graduation. Their findings suggest that the reform in fact reduced labor market entry age of the G8 cohort. Regarding potential adverse reform effects, they find that women affected by the reform tend to start their career with a lower degree of education. They do not find significant effects on job matching quality as reflected in the adequacy of required qualifications. However, the survey data bases of the Saxony-Anhalt studies has been criticized to suffer from selectivity (Homuth, 2017, p. 43).[11]

As for students in the academic track schools, we assume that the reform yields higher distress. Increasing the hours spent in school, the reform directly affects students' workload

affected aesthetic subjects (arts, music, sports). This total reduction amounts to about three percent of the overall time spent in school. However, Homuth (2017) also states that the observed hours are the minimum of required hours; it is likely that students on average actually spend some more time in schools. While the illustrated features apply to the entity of the federal states implementing the reform, precise curricular changes somewhat differ on the state level.

[11]Concerning transition to tertiary education, Meyer et al. (2016) analyze German secondary education graduate survey data. Based on a difference-in-difference strategy using state and time variation in the reform, they find patterns of delayed entrance to universities. Effects are significant for female students who more often participate in voluntary services or travel abroad after high school (*Gymnasium*). Further findings for delayed university entrance of double cohort students in the German state of Saxony-Anhalt are provided in Büttner and Thomsen (2013) and Meyer and Thomsen (2016).

as well as their time budget. In our empirical analysis we will address leisure time effects and related channels through which the reform affects students' life.[12]

To date, the existing literature drawing on the difference-in-difference identification strategy tends to report positive performance effects for students while observed in the academic track. Based on 2000-2009 PISA data, Andrietti (2015) finds no significant effects of the reform on grade retention but positive effects on students' test scores. Similarly, Huebener et al. (2017) demonstrate higher test scores of 15 year old G8 students. Quantile regression results point to smaller effects for students at the bottom of the performance distribution. Similar findings are reported in Homuth (2017).[13] However, based on administrative data, Huebener and Marcus (2015a) suggest that the incidence of grade repetition increases (especially for boys) which implies that the reform decreases the average schooling years by less than one year. The study finds no general impact on secondary academic track graduation rates. Further studies refer to comparisons within federal states: Büttner and Thomsen (2013) find negative effects on math grades at graduation. However, the latter result refers to the specific group of students in the double cohort of the state of Saxony-Anhalt. Kühn et al. (2013) summarize further (correlative) evidence which does not point to systematic performance differences. Dörsam and Lauber (2015) observe no performance differences for university students at one specific German university (Konstanz).

3 Empirical Strategy: Using the G8 academic track reform in Germany as a natural experiment

Our aim is to further explore the relationship of skill investment in the school context on the formation of (mental) health in adolescence. In line with the public perception of the G8 reform, we expect the reform to be a negative investment in the formation of (mental) health as it increases distress in school. To explore such a relationship between the G8 reform and specific health outcomes, we first consider the following regression equation:

$$H_{ist} = \alpha\, R_{ist} + f(age_{ist}) + \xi_{ist} \tag{1}$$

where H_{ist} is the respective health measure of individual i in state s from cohort t, R_{ist} is the reform indicator, and $f(age_{ist})$ is a polynomial function of age which controls for

---

[12]In the previous literature, Meyer and Thomsen (2015) investigate workloads and leisure time activities of double cohort graduating students in the state of Saxony-Anhalt. Based on survey data for 12 schools, the study finds that G8 students more often report to suffer from high workloads. Along with this, G8 students more often attend tutoring lessons, less often participate in volunteering activities or part-time work and have less time for relaxing activities such as watching TV, listening to music or reading. Correlative evidence reported in Trautwein et al. (2015) mostly points to some differences but overall similar leisure time use of students in G8 and G9 schools in Baden-Württemberg.

[13]Findings may be biased because of inclusion of potentially endogenous control variables (e.g. grade repetition) in the regression analysis.

general age trends in health.[14] The coefficient α indicates the general correlative relationship between the reform and the health measure. Note that α does not imply a causal effect. Specifically, we assume that there are cohort trends in health.[15] A negative mental health cohort trend in this model would imply that G8 students more often suffer from poor mental health even in the absence of a reform effect – just because they are more often observed in a more recent school entry cohort in our sample. Omitted variables (e.g. cohorts) will be included in the error term ξ_{ist} of equation 1. Correlation of the error term and the reform indicator will cause a biased estimate of α. Similarly, there may be general baseline effects of federal states that will bias the impact estimates.

Starting from equation 1, we consider two further correlative specifications, where we add federal state indicators (specification Ia) and school entry cohorts (Ib) respectively. In order to estimate causal effects of the G8 reform, we exploit the fact that the reform was implemented at different points in time across federal states. Whether a student was affected by the reform or not, depends on the year of school entry as well as his or her state of residence at that point of time. We estimate the following difference-in-difference model:

$$H_{ist} = \beta\, R_{ist} + f(age_{ist}) + \sum_{s=1}^{S} \mu^s state_{ist}^s + \sum_{y=1}^{T} \pi^t year_{ist}^t + \zeta_{ist} \qquad (2)$$

where $state_{ist}^s$ refers to a full set of federal state indicators and $year_{ist}^t$ to a full set of observed school entry cohorts. Thus, the coefficient of main interest, β, identifies the general G8 reform effect within state and cohort. In the quasi-experimental setting of our analysis, we assume that differences in individual background variables are not systematically affecting the observed health outcomes. However, we still include a set of individual characteristics, X_{ist} to inform on a third specification:

$$H_{ist} = \gamma\, R_{ist} + f(age_{ist}) + X_{ist}\theta + \sum_{s=1}^{S} \delta^s state_{ist}^s + \sum_{t=1}^{T} \lambda^t year_{ist}^t + \epsilon_{ist} \qquad (3)$$

Here, γ identifies the reform effect. The control variables include individual and family background characteristics: Students' gender, an immigrant background indicator (for either the student or at least one parent having been born abroad), parental education (i.e. an indicator for at least one parent holding a higher secondary degree), an indicator for maternal employment, indicators for being a single child and for being a first born child, an indicator for single-headed families and an indicator for living on the countryside. Table A-1 in the Appendix contains definitions of all background variables. Balancing tests on these background variables do not point to any relevant differences between G9 and G8 cohorts that could pose a threat to our estimation strategy after controlling for age, federal state and school entry cohort. Table A-2 in the Appendix presents the results for these tests.

---

[14]The results we report in the following sections refer to a second order polynomial of age. We have also tested third order polynomials. However, the polynomial term coefficients are not statistically significant.

[15]We observe individuals entering school in different cohorts. Because observations stem from different survey waves, age does not yield the same variation as the cohort information in our data.

We generally report robust standard errors clustered on the federal state level because this is the level relevant for education policy variation (following Cameron and Miller, 2015).

There are several key assumptions that are relevant for our strategy to identify meaningful causal effects of the G8 reform. We have to assume that in the absence of the reform the evolution of students' health and health behavior would follow the same trend over time in all federal states, irrespective of the initial average health level in each state. This common trend assumption is generally not testable. We see, however, no evidence that could disprove this assumption. To this end, Huebener et al. (2017) show that other major school reforms such as the introduction of central exit examinations and changes in the point of time of tracking after primary school do not coincide with the introduction of G8.

To challenge our findings, we provide insight from a randomization inference exercise based on random assignment of reform years to federal states. Among our robustness checks, we also perform placebo regressions with students only from the vocational secondary tracks who were not directly affected by the G8 reform. In case of other influencing (policy) factors that coincide with the G8 reform, we should find significant effects for these students as well. However, there may also be indirect effects of the reform on students in vocational track schools. Particularly, due to a higher number of academic track school graduates in the graduation year of the first G8 cohort (double cohorts), vocational school graduates face a tougher competition for vocational training positions. Note that academic as well as vocational track students compete for the "better" vocational positions, since not all of the academic track students opt for tertiary (university) education. Harder competition could in turn influence vocational track students' (mental) health. Thus, even if the zero effects we find for the placebo group bolster confidence in our empirical design, we are aware of the vocational secondary track schools to be a somewhat imperfect placebo sample. Therefore, we do not draw on this control group in a triple difference design as our main identification strategy.[16]

Furthermore, we assume that whether a student is affected by the reform or not is not correlated to students' unobserved characteristics also affecting the outcome measures – that is treatment depends only on date of birth and federal state of residence, but not on the student's deliberate choice. Theoretically, selection into or out of treatment may occur if students move to a different federal state that has not yet (or not already) implemented the reform. However, as documented in Huebener et al. (2017), commuting or moving to a different federal state is unlikely as it implies high costs for the family. Another possibility to avoid attending the secondary school level under G8 would be to opt for an alternative school track such as *Realschule* or *Hauptschule*. As both school types, however, do not yield a university entrance qualification but instead generally prepare for a vocational education, opting for these tracks can be considered to be rather unlikely. This assumption is in line with findings in Huebener and Marcus (2015b) who do not find evidence for lower enrolment rates

---

[16]In fact, while most previous (causal) studies follow a difference-in-difference strategy, (Westermaier, 2016) is an exception drawing on a triple difference strategy including vocational track students.

at academic track schools induced by the G8 reform using administrative data on students in Germany.

In addition, we perform own analyses of the effect of the G8 reform on the share of students in academic track schools in Germany using the SOEP data. As participants in the SOEP are generally interviewed for the first time at age 17, there is no possibility to analyze the direct transition of students into secondary schools in our data. We therefore look at the share of students in academic track schools at the age of 17.[17] Our results confirm the finding in Huebener and Marcus (2015b) that the reform did not have a significant impact on the general school choices of students (results available upon request).[18]

4 Data

Our study is based on data of the German Socio-Economic Panel (SOEP), a representative household panel survey, covering the years 1984 to 2014 and including about 20,500 individuals in the most recent wave (Glemser et al., 2014). For our analysis, we use a pooled cross-section of the years 2005 to 2014, covering the education reform period. The SOEP provides us with variables on health outcomes and health behavior as well as with a rich set of information on education, family background and socio-economic status.

In order to identify whether a student was affected by the G8 reform or not, we use information on the year of school entry, the federal state of school attendance and the type of school visited. Since we do not observe the information on schooling year for all students, we allocate students to school cohorts according to their date of birth and school entry cut-off dates.[19] In our main analysis, we only include students who attended an academic track school throughout their secondary schooling time. Only these students have been affected by the reform.[20] Our sample consists of 1,130 individuals in academic track schools, of whom 415 students (36.7 %) are affected by the G8 reform.

Survey participants are generally interviewed for the first time in the year they turn 18 and followed up once in a year.[21] As questions on both health outcomes and health behavior are asked only every second year, participants possibly answer the same questions for the first time at different ages. In our sample, we therefore observe every person either in the year they turn 18 or in the year they turn 19. Furthermore, there are some individuals entering

---

[17]At this age, academic track school students are still attending school, irrespective of whether they are affected by the G8 reform or not.

[18]Besides, if the G8 reform exerted an impact on the composition of the students within school track, resulting health effects should also be reflected in vocational track students. We test this assumption with our placebo regressions.

[19]Information for the applicable cut-off dates stems from the national education report (Autorengruppe Bildungsberichterstattung, 2012).

[20]Only in our placebo regression, we draw on students in the more vocationally oriented secondary track schools who are not affected by the reform.

[21]Since 2006, adolescents aged 16 to 17 are being surveyed by means of special youth questionnaires. However, only the regular questionnaire for adults contains the detailed questions on self-assessed health that we use for our analysis.

the panel study when they are somewhat older. In light of the limited sample size, we include these later panel entrants as well, but restrict our sample to young adults no older than 25.[22] Table A-3 in the Appendix presents the corresponding numbers of survey participants by age group in the G8 and G9 cohorts. The table also shows the numbers of G8 and G9 students by federal state. Our sample includes students from the federal states of Hamburg, Lower Saxony, Northrhine-Westphalia, Baden-Württemberg, Bavaria, Berlin, Mecklenburg-Western Pomerania and Saxony-Anhalt. These are the states switching from the G9 to the G8 regime at a specified year within our period of observation.[23]

Health measures used in this study relate to two standard summary scales on general mental and physical health, i.e. the Mental Component Summary Scale (MCS) and the Physical Component Summary Scale (PCS). We also examine the eight sub-scales available in the SOEP. Measurement of the MCS and PCS is based on a slightly modified version of the well-established Short Form Health Survey (SF-12v2), which is a 12-item subset of the SF-36v2 and has been included in the SOEP since 2002 (Andersen et al., 2007; Schupp and Wagner, 2007). Table A-5 in the Appendix depicts the original questions from the SOEP questionnaire as well as the corresponding sub-scales. We present standardized scores with mean 50 and standard deviation 10 (cf. Andersen et al., 2007). The MCS and PCS are objective and comprehensive health measures (Ziebarth, 2010): While the PCS scale measures overall health (Andersen et al., 2007), the MCS scale has been shown to be a valid indicator of mental disorders (Vilagut et al., 2013; Gill et al., 2007). In order to judge the size of our estimates, one may relate to findings from studies on specific patient populations: For example, according to Vilagut et al. (2013), the cutoff for 30-day episodes of depressive disorders is about 44 % of a standard deviation below the mean 50. For physical outcomes, for example Müller-Nordhorn et al. (2004) show that the PCS calculated from the SF-12 in a German sample of patients suffering from coronary heart disease is reduced about one standard deviation below the mean 50. We consider such reference effect sizes in the concluding discussion of our results (Section 7).

In addition to the MCS and PCS, we address measures of health behavior. To this end, we define indicators for overall tobacco consumption as well as for regular (>5 cigarettes per day) and for heavy (>15 cigarettes per day) tobacco consumption. The underlying information on smoking behavior is collected once every two years.

Table 1 reports descriptive statistics on health outcomes and on background characteristics for students in the G8 and G9 school schemes. Mean values of health outcomes are similar for students under both regimes. Regarding health behavior, however, G8 students tend to smoke less than students in the G9 scheme.

---

[22]Our sub-group analyses address potential heterogeneous effects depending on age and the corresponding stage in the education context.

[23]In Hesse, schools could switch to the new system within a three year transition time. We do not include Hesse in our sample, because we cannot clearly distinguish between treatment and control status. Furthermore, three smaller states in terms of population size (Brandenburg, Bremen and Saarland) are not in our sample as we do not observe students from cohorts under the reform in our data.

Table 1: Summary statistics for health outcomes and background variables

	G9 students			G8 students		
	Obs.	Mean	(Sd)	*Obs.*	Mean	(Sd)
Health outcomes						
MCS Score	714	50.13	(9.59)	403	49.07	(9.50)
– Vitality	714	53.28	(9.25	403	51.92	(9.76)
– Social interaction	714	52.37	(8.22)	403	52.28	(8.46)
– Work ability (mental)	714	52.10	(9.33)	403	51.61	(9.92)
– Emotional balance	714	50.87	(9.18)	403	49.34	(8.90)
PCS Score	714	57.38	(5.26)	403	56.95	(6.67)
– General health	714	56.76	(7.81)	403	56.56	(8.55)
– Functional ability	714	56.85	(3.94)	403	56.19	(5.27)
– Work ability (physical)	714	55.43	(6.70)	403	54.14	(8.09)
– Pain	714	54.64	(7.12)	403	53.97	(7.48)
Smoking (yes/no)	704	0.18	(0.38)	401	0.14	(0.35)
Smoking (>5 cig. per day)	704	0.12	(0.33)	401	0.09	(0.29)
Smoking (>15 cig. per day)	704	0.04	(0.20)	401	0.02	(0.14)
Background variables						
Age	716	19.62	(1.65)	415	18.70	(0.91)
Male	716	0.47	(0.50)	415	0.45	(0.50)
Migration background	716	0.18	(0.39)	415	0.24	(0.43)
High parental education	716	0.63	(0.48)	415	0.57	(0.50)
Non-working mother	716	0.11	(0.32)	415	0.13	(0.34)
Single child	716	0.11	(0.32)	415	0.11	(0.31)
First born child	716	0.55	(0.50)	415	0.64	(0.48)
Single parent	716	0.18	(0.39)	415	0.27	(0.44)
Rural	716	0.22	(0.41)	415	0.29	(0.46)

Note: Summary statistics based on SOEP waves v23-v31 (2006-2014).

With respect to background characteristics, Table 1 reveals some differences between the G8 and G9 cohorts: The G8 students we observe in our sample are on average about a year younger than students observed under the longer school regime (G9). Observed G8 students somewhat more often have an immigrant background, have parents with lower education levels and live in single headed households and in rural areas. These differences depict two issues: First, more recent cohorts appear more often in the G8 sample. Secondly, there is

heterogeneity over the respective federal states. In our identification strategy, we assume that there are no systematic outcome-related patterns of states entering the reform earlier or later. As stated in Section 3, balancing tests on background variables support the validity of this assumption (see Table A-2 in the Appendix).

5   Results, sub-group analyses and mechanisms

Table 2 provides the G8 reform coefficient estimates for regressions of the MCS and PCS scores and sub-scales (top panels) or the alternative tobacco consumption indicators (bottom panels). The left part of Table 2 contains results from the correlative specifications regressing health outcomes on the reform indicator and age (I) and adding federal state indicators (Ia) or school entry cohort (Ib), respectively. Our main results are provided on the right columns of Table 2, where specifications II and III draw on the difference-in-difference strategy (cf. Section 3). Specifications II and III relate to our central identification strategy; we will more carefully discuss the corresponding results in the following. Generally, the correlative evidence is rather robust between specifications I and Ia, while adding cohort indicators and thus eliminating time trends in specification II yields estimates that tend to be closer to the difference-in-difference-findings (II and III).

All point estimates on the PCS and MCS scores as well as the available sub-scales are consistently negative. For the MCS score, the point estimates are statistically significant and robust across the difference-in-difference specifications (II and III), implying that G8 students suffer from lower self-assessed mental health: According to specification II (III), ceteris paribus the MCS score is about -2.73 (-2.71) points lower than for G9 students. This effect corresponds to about one quarter of an MCS standard deviation. The MCS sub-scale coefficients reveal that this negative finding is significantly driven by lower self-perceived vitality and less favorable emotional balance (with all point estimates being negative for the sub-scale effects).

With respect to physical health, the coefficient on the overall PCS score is negative but not statistically significant. However, in the difference-in-difference specifications, we find a significant and robust effect on the sub-scale of general self-assessed health. According to these estimates, the G8 reform reduces self-rated general health by about one quarter of a standard deviation (-2.55). We consider this result crucial as precisely this simple measure of self-rated general health is a commonly used health indicator.

Table 2: Effect of G8 reform on health outcomes

| | Coefficient on G8 reform | | | | |
| | Correlative | | | Diff.-in-Diff. | |
Health measure	(I)	(Ia)	(Ib)	(II)	(III)
MCS Score	-1.128 *	-0.769	-3.274 **	-2.728 **	-2.705 *
	(0.616)	(0.546)	(1.054)	(1.109)	(1.219)
– Vitality	-1.141 *	-1.004	-3.225 **	-3.549 ***	-3.499 ***
	(0.624)	(0.837)	(1.023)	(0.913)	(0.997)
– Social Interaction	-0.343	-0.174	-1.103	-0.503	-0.460
	(0.542)	(0.356)	(0.881)	(0.892)	(0.889)
– Work ability (mental)	-0.733	-0.433	-2.694	-2.329	-2.249
	(0.633)	(0.738)	(1.442)	(1.484)	(1.540)
– Emotional Balance	-1.616 ***	-1.282 ***	-3.547 **	-2.956 **	-3.047 **
	(0.585)	(0.325)	(1.089)	(0.977)	(1.141)
PCS Score	-0.648	-0.769 *	-0.679	-0.943	-0.941
	(0.401)	(0.347)	(1.153)	(1.308)	(1.274)
– General Health	-0.399	-0.372	-2.219	-2.554 *	-2.554 *
	(0.546)	(0.400)	(1.229)	(1.344)	(1.204)
– Functional ability	-0.952 ***	-0.898 ***	-1.658 **	-1.373	-1.415
	(0.303)	(0.230)	(0.596)	(0.753)	(0.759)
– Work ability (physical)	-1.547 ***	-1.568 **	-1.433	-1.307	-1.229
	(0.498)	(0.564)	(1.277)	(1.566)	(1.514)
– Pain	-0.711	-0.755	-0.693	-0.979	-1.011
	(0.485)	(0.535)	(1.284)	(1.391)	(1.393)
N	1, 117	1, 117	1, 117	1, 117	1, 117
Smoking, (yes/no)	-0.018	-0.014	0.004	0.003	0.003
	(0.023)	(0.022)	(0.038)	(0.036)	(0.037)
Smoking, >5 cig. per day	-0.018	-0.012	-0.004	0.017	0.021
	(0.019)	(0.018)	(0.023)	(0.023)	(0.023)
Smoking, >15 cig. per day	-0.015	-0.012	-0.007	0.011	0.015
	(0.010)	(0.010)	(0.005)	(0.010)	(0.013)
N	1, 105	1, 105	1, 105	1, 105	1, 105
Fed. state indicators	no	yes	no	yes	yes
Cohort indicators	no	no	yes	yes	yes
Control variables					
– Age	yes	yes	yes	yes	yes
– Backgr. characteristics	no	no	no	no	yes

Note: Results based on observations from SOEP waves v23-v31 (2006-2014). Specifications: (I) Baseline specification: OLS regression of health outcome on treatment (G8 participation) including age and age squared as controls; (Ia) Basline specification including federal state indicators; (Ib) Baseline specification including school entry cohort indicators; (II) Baseline specification including both federal state *and* school entry cohort indicators; (III) Baseline specification including federal state and school entry cohort indicators and including the full set of background variables as controls. Cluster-robust standard errors in parentheses, * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

For our health behavior indicators (smoking), the point estimates in the basic specification (I and Ia) suggest a negative correlation between being a G8 student and smoking. Part of this pattern is driven by a general negative cohort trend in smoking, as implied by the lower estimates in specification Ib. However, the point estimates turn consistently positive in the difference-in-difference specifications (II and III). While these coefficients are not statistically different from zero, the coefficients tentatively hint to G8 students more often adopting negative health behaviors (related to tobacco consumption). This pattern will be confirmed in some of the sub-group analyses below, where we find statistically significant estimates.

We may question whether the effects identified in Table 2 are true reform effects or rather due to initial problems related to the periods of transition to the new regime. For example, the effects may be driven by the specific situation in the first G8 year (double cohort, cf. the above Section on institutional background). Similarly, one may hypothesize that higher learning intensity exerts a mere temporary impact on (mental) health and that this impact vanishes after graduation. As we observe adolescents and young adults aged 18 to 25 from several school entry cohorts both before and after the implementation of the reform, we are able to distinguish between different cohorts and educational subgroups. These results are presented in Table 3. Columns A1 and A2 of Table 3 contain separate results for the transition year to the reform (double cohorts) and regular G8 attendants. Columns B1 and B2 extend this view and present results relating to "early cohorts" that are affected by the reform in the first two years and "late cohorts" (at least three years after implementation of the reform). Results from these four sub-samples do not point to the overall effects being solely driven by cohorts in the transition period. On the contrary, for the mental health sub-scales, the coefficients which are still significant in the reduced samples tend to be more pronounced for the more recent cohorts (column A2 and B2 as compared to A1 and B1, respectively). On the other hand, in the transition samples, we observe negative reform effects on functional ability (as well as a significant higher incidence of being a regular smoker in the double cohort sample). This may point to a specific transition time effect – also reflected in the overall PCS score – on health. However, as we do not observe such an effect in our overall sample 2, this should not affect our general conclusions on the remaining outcomes.

In columns C1 and C2 of Table 3 we consider current high school students and high school graduates, respectively. We do not observe a systematic pattern of lower coefficients on mental health outcomes for high school graduates (C2) as compared to current academic track students (C1). Even if the aggregated MCS score is insignificant for graduates, significant point estimates related to some of the MCS sub-scales are even higher in this group (vitality, emotional balance). None of these differences in coefficients from the two samples (C1 and C2) are statistically significant. If effects vanish over time, after leaving school, this seems to be mainly related to physical capacities (general health, functional ability) as well as to mental work abilities, as for these outcomes the differences are in fact statistically significant.

17

Table 3: Effect of G8 reform on health outcomes for student subgroups

Health measure	Cohort		Implementation stage [a]		Student status	
	Double cohort (A1)	Regular cohort (A2)	Early cohorts (B1)	Late cohorts (B2)	Student (C1)	Graduate (C2)
MCS Score	-2.250	-3.848	-2.098	-3.158**	-3.327*	-2.345
	(2.538)	(2.254)	(2.826)	(0.937)	(1.730)	(2.010)
– Vitality	-3.608	-4.718**	-3.825	-4.913***	-3.217*	-4.886**
	(3.382)	(1.564)	(2.055)	(1.122)	(1.396)	(2.011)
– Social Interaction	-1.092	-0.321	-0.715	-0.713	-0.717	-0.308
	(2.318)	(0.969)	(2.407)	(2.711)	(0.823)	(1.559)
– Work ability (ment.)	-2.093	-2.766	-3.939	-1.779	-4.537**	1.805
	(3.415)	(2.006)	(2.482)	(1.933)	(1.623)	(1.487)
– Emotional Balance	-3.543*	-4.324*	-1.713	-2.739***	-3.243	-3.252*
	(1.511)	(2.188)	(2.038)	(0.743)	(1.761)	(1.474)
PCS Score	-2.794	-0.062	-2.902*	0.469	-1.749	0.750
	(1.943)	(1.627)	(1.351)	(2.311)	(1.185)	(1.248)
– General Health	-3.220	-1.442	-3.283	-0.166	-4.119**	1.561
	(4.053)	(0.821)	(2.020)	(1.829)	(1.198)	(1.831)
– Functional ability	-4.551**	-2.320	-3.717*	-1.282	-2.119*	0.181
	(1.592)	(1.327)	(1.724)	(1.612)	(0.973)	(0.501)
– Work ability (phys.)	-1.244	-0.897	-1.273	0.801	-2.016	-1.235
	(2.608)	(1.172)	(2.084)	(1.605)	(1.634)	(1.344)
– Pain	-1.244	-0.897	-1.273	0.801	-2.016	-1.235
	(2.608)	(1.172)	(2.084)	(1.605)	(1.634)	(1.344)
N	158	959	272	845	734	383
Smoking, (yes/no)	0.130	0.070	0.059	0.158	-0.058	0.105
	(0.089)	(0.080)	(0.052)	(0.115)	(0.074)	(0.057)
Smoking, (>5 cig. per day)	0.178*	0.084	0.068	0.092*	-0.028	0.086
	(0.086)	(0.063)	(0.048)	(0.040)	(0.065)	(0.046)
Smoking, (>15 cig. per day)	0.049	0.026	0.025	0.027	0.000	0.025
	(0.052)	(0.019)	(0.023)	(0.016)	(0.018)	(0.015)
N	149	956	258	847	662	443
Fed. state indicators	yes	yes	yes	yes	yes	yes
Cohort indicators	yes	yes	yes	yes	yes	yes
Control variables						
– Age	yes	yes	yes	yes	yes	yes
– Backgr. charact.	yes	yes	yes	yes	yes	yes

[a] The implementation stage refers to whether the graduation cohorts were affected by the reform as one of the first cohorts. "Early" cohorts = Double cohort plus the second cohort affected by the G8 reform.
Note: Results based on observations from SOEP waves v23-v31 (2006-2014). OLS regression of health outcome on treatment (G8 participation) including age and age squared, both federal state and school entry cohort indicators and the full set of background variables as controls. Cluster-robust standard errors in parentheses, * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

In sum, from these subgroup results we conclude that the observed patterns do not generally point to the G8 reform solely affecting students in the early years after reform implementation. We also conclude that the (mental) health effects remain sizeable after students graduate from the higher academic track.

As a further hypothesis, one may expect that the mental health impact is particularly pronounced for more 'vulnerable' groups, for example for students at the bottom of the health distributions or those with less favorable home resources. In the following, we shed further light on such heterogeneous effects. We conduct unconditional quantile regressions related to the MCS and PCS distributions, respectively. In the previous literature, Huebener et al. (2017) conduct quantile regressions looking at heterogeneous performance effects of the G8 reform. In close analogy to their paper, we consider unconditional quantiles, where we apply the recentered influence function (RIF, following Firpo et al., 2009). The corresponding results are presented in Figure 2.[24] We do not find a consistent pattern over the considered deciles. As for mental health, there seem to be mostly homogeneous reform effects for students in all deciles of the MCS distribution. With respect to physical health, we observe significant negative effects for students in the lower PCS deciles. These results indicate that particularly students of below average physical health are negatively affected by the reform.

Figure 2: Quantile regressions by deciles of the MCS and PCS distributions

Note: Estimates based on SOEP waves v23-v31 (2006-2014). RIF-regression estimates of the reform effect on MCS and PCS scores; regressions include the full set of control variables. Confidence bounds are calculated based on bootstrapped standard errors (50 replications).

To further address potentially vulnerable sub-groups, we conduct regressions including interaction terms between the reform indicators and each of the observed background variables. Most of the resulting coefficients are insignificant (including the baseline G8 effects). For the MCS results, we still obtain significant negative G8 effects for students being a single child and for students not being a first born child. When considering the health behavior outcomes, we find that the reform seems to specifically increase tobacco consumption among

---

[24]Standard errors are bootstrapped. We do not report results for the MCS and PCS sub-scales. For most of these sub-scales, we do not observe sufficient heterogeneity in order to obtain meaningful results.

students having an immigrant background. For heavy tobacco consumption of more than 15 cigarettes a day, we find that G8 students more often adopt this behavior if their mothers are not working or if none of their parents holds a higher secondary education degree. Results for the sub-group regressions are shown in Table A-4 in the Appendix.[25] In summary, the sub-groups' results tentatively point to socially disadvantaged students being somewhat stronger affected by negative effects of the reform.

In order to gain further insight into the reform effects, we examine a set of outcome variables related to student workload and students' daily life. For the workload dimension, we observe perceived (self-reported) time pressure. We also draw on information regarding whether students attend any form of private tutoring classes. For further daily life considerations, we observe whether the student (at age 15) often argued with parents. We also examine students' extracurricular activities (weekly sport and music activities, daily television consumption and use of computer games) as well as their relationships with friends (information on weekly meetings with friends or peers). The G8 effects on these outcomes are estimated based on the same set of control variables as in the health regressions. Results are presented in Table 4 based on the affected sample of academic track students.

Regarding the workload dimension, we find positive (but marginally insignificant) point estimates for the G8 reform's effect on time pressure. We also find a significant reform effect on the probability of attending private tutoring. Students under the G8 reform are about 14 percentage points more likely to attend tutoring classes. However, with respect to the daily life outcomes, Table 4 shows that most of the coefficients are not statistically significant from zero. We still find that the G8 reform significantly increases the probability of frequent disputes with parents by about 11 percentage points. Our results are in line with findings in Meyer and Thomsen (2015) who also report a higher perceived learning pressure which is accompanied by an increase in private tutoring.

---

[25]We do not report results for the MCS and PCS sub-scales. These results are available upon request and show a pattern consistent to the provided evidence.

Table 4: Effect of G8 reform on extracurricular activities, time use and relationship with parents

Activity	N	Mean	Correlative			Diff.-in-Diff.	
			(I)	(Ia)	(Ib)	(II)	(III)
Time pressure	746	0.06	0.017	0.014	0.055	0.058	0.055
			(0.017)	(0.013)	(0.045)	(0.040)	(0.045)
Private tutoring	747	0.32	0.036	0.056 **	0.032	0.136	0.143 *
			(0.036)	(0.021)	(0.080)	(0.072)	(0.074)
Dispute w/ parents	747	0.30	0.023	0.027	0.063	0.112 **	0.111 **
			(0.035)	(0.040)	(0.062)	(0.041)	(0.041)
Sports activities	610	0.84	0.080 ***	0.089	0.043	0.073	0.067
			(0.030)	(0.055)	(0.075)	(0.081)	(0.088)
Music activities	605	0.44	-0.010	-0.008	-0.042	-0.036	-0.052
			(0.044)	(0.048)	(0.076)	(0.060)	(0.059)
Watching TV	609	0.69	-0.124 ***	-0.120 ***	-0.041	0.016	0.025
			(0.041)	(0.030)	(0.052)	(0.064)	(0.059)
PC games	609	0.27	-0.040	-0.029	-0.030	0.000	-0.013
			(0.038)	(0.052)	(0.062)	(0.065)	(0.049)
Meeting friends	605	0.82	-0.010	-0.017	-0.004	0.015	0.022
			(0.034)	(0.046)	(0.054)	(0.113)	(0.114)
Fed. state indicators			no	yes	no	yes	yes
Cohort indicators			no	no	yes	yes	yes
Control variables							
– Age			yes	yes	yes	yes	yes
– Backgr. characteristics			no	no	no	no	yes

Note: Results based on observations from SOEP waves v23-v31 (2006-2014). Specifications: (I) Baseline specification: OLS regression of treatment (G8 reform) on time use and relationship outcomes including age and age squared as controls; (Ia) Baseline specification including federal state indicators; (Ib) Baseline specification including school entry cohort indicators; (II) Baseline specification including both federal state *and* school entry cohort indicators; (III) Baseline specification including federal state and school entry cohort indicators and including the full set of background variables as controls. Cluster-robust standard errros in parentheses, * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

6 Sensitivity analysis and placebo regressions

As indicated in Section 3, our identification strategy and the interpretation of the findings being representative effects relies on a variety of assumptions which we will tackle in the following. An overview of these sensitivity analyses is given in Table 5. The first column of Table 5 (III) repeats our full specification regression outcomes. The analyses underlying columns A-C challenge our estimation strategy in that they consider whether any of the assumptions set out in Section 3 do not hold by considering specific subsamples.

Theoretically, even if considered unlikely, students may select out of the G8 scheme through entering secondary school in a different state where the reform has not yet been implemented. Opting out may bias our estimates towards zero if we expect that avoiders consist of the group of students that fears to be most negatively affected by the reform, for example as they lack other resources to cope with distress in school. This strategy, however, only works up to the point of time when all federal states have implemented the reform. Therefore, column A of Table 5 is based on a sample only including late-adopting states (Baden-Württemberg, North Rhine-Westphalia and Berlin), where students did not have the possibility to avoid G8 by moving to another state. While the general findings on mental and physical health are robust, many of the estimated effects (except of the indicator on emotional balance) do become somewhat more pronounced based on an examination of point estimates. Additionally, we obtain significant (positive) impacts on the smoking indicator in this sample.

A second possibility to avoid the G8 scheme would be to opt for an alternative school track, either a comprehensive secondary school *Gesamtschule* or a vocationally oriented secondary school (see Section 2).[26] In column B, the sample only includes states where comprehensive schools do typically not exist (Baden-Württemberg, Bavaria, Mecklenburg-West Pomerania and Saxony-Anhalt). Again, the results are robust. However, there is a tendency of more pronounced (but still robust) MCS point estimates in this sample. In sum, the MCS and PCS point estimates in both samples A and B – being mostly robust in statistical terms – may point to some degree of selectivity biasing the results towards zero in the overall sample.

Column C of Table 5 relates to our most challenging robustness check addressing the common trend assumption. In this sample, we only include students in vocational secondary schools. As these are shorter tracks per se (about 5-6 years of secondary schooling), students from these schools should not be directly affected by the reform. If we found significant coefficients in this sample, this could challenge our main regression results (or point to indirect effects of the reform on vocational track students, cf. Section 3). Even if this control group might not be perfect, the coefficients in Table 5 again bolster confidence in our identification strategy with all of the coefficients being insignificant in the placebo regression sample C.

These placebo results indicate that there do not seem to be significant health effects of the G8 reform on non-academic track students. Still, as noted above, we cannot completely rule out selection effects. Therefore, we do not draw on a triple difference strategy as our main identification strategy. However, we have also estimated triple difference regressions based on our data. The general results are robust, with insignificant results for the additional interaction effect of the vocational track school indicators and the reform coefficient.

---

[26]In comprehensive schools, students may still obtain the university entrance qualification after 9 years of secondary education.

Table 5: Sensitivity analysis for main results

Health measure	Overall sample (III)	Late-adopting states (A)	No states w/ Gesamtschule (B)	Stud. in lower sec. track (placebo) (C)
MCS Score	-2.705 *	-3.888	-4.901 **	-0.088
	(1.219)	(2.317)	(1.459)	(1.773)
– Vitality	-3.499 ***	-4.384 *	-5.181*	0.902
	(0.997)	(1.200)	(1.845)	(2.446)
– Social Interaction	-0.460	-2.296	-0.063	-0.029
	(0.889)	(1.867)	(1.167)	(1.302)
– Work ability (mental)	-2.249	-4.854 *	-3.340	-0.378
	(1.540)	(1.523)	(2.550)	(0.727)
– Emotional Balance	-3.047 **	-2.588	-5.862 **	-0.498
	(1.141)	(1.780)	(1.451)	(1.813)
PCS Score	-0.941	-2.105	0.073	-0.002
	(1.274)	(1.275)	(2.297)	(0.661)
– General Health	-2.554 *	-3.597	-2.486	0.767
	(1.204)	(1.434)	(2.167)	(1.670)
– Functional ability	-1.415	-2.072	-0.929	-0.155
	(0.759)	(0.722)	(1.041)	(0.876)
– Work ability (physical)	-1.229	-2.832	-1.460	0.307
	(1.514)	(1.610)	(2.532)	(1.139)
– Pain	-1.011	-0.995	-2.329	-1.224
	(1.393)	(2.677)	(1.079)	(1.315)
N	1,117	521	614	1,003
Smoking, (yes/no)	0.003	0.114 *	-0.048	0.092
	(0.037)	(0.036) *	(0.037)	(0.053)
Smoking, >5 cig. per day	0.021	0.082 ***	-0.024	0.108
	(0.023)	(0.008)	(0.029)	(0.059)
Smoking, >15 cig. per day	0.015	-0.006	0.055	-0.005
	(0.013)	(0.012)	(0.025)	(0.056)
N	1,105	517	602	941
Fed. state indicators	yes	yes	yes	yes
Cohort indicators	yes	yes	yes	yes
Control variables				
– Age	yes	yes	yes	yes
– Backgr. characteristics	yes	yes	yes	yes

Note: Results based on observations from SOEP waves v23-v31 (2006-2014). Specifications: (I) Main specification: OLS regression of health outcome on treatment (G8 participation) including age and age squared, both federal state and school entry cohort indicators and including the full set of background variables as controls; (II) Main specification, including only "late-adopter" states (Baden-Württemberg, North Rhine-Westphalia and Berlin; (III) Main specification, including only states where comprehensive schools do typically not exist (Baden-Württemberg, Bavaria, Mecklenburg-West Pomerania and Saxony-Anhalt); (IV) Main specification, including only students who attended a lower secondary school track and were thus not affected by the G8 reform. Cluster-robust standard errors in parentheses, * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

In order to address the significance of our results, we provide randomization inference based on randomly assigning students to the G8 reform (following Chetty et al., 2009). We conduct 6,000 placebo regressions based on random draws of treatment-year-state combinations, i.e. we assign random numbers to states and order them according to these numbers to consecutively assign reform years. We then estimate the reform effect based on this treatment assignment. Figure 3 plots the distributions of the resulting reform coefficients on MCS, PCS and the overall tobacco consumption indicator together with lines referring to the initially estimated treatment effects and p-values (indicated in boxes). Again, Figure 3 clearly points to statistical significance of the mental health effect where we do not expect that this is a randomly obtained result.

Figure 3: Placebo Regressions MCS, PCS and Smoking
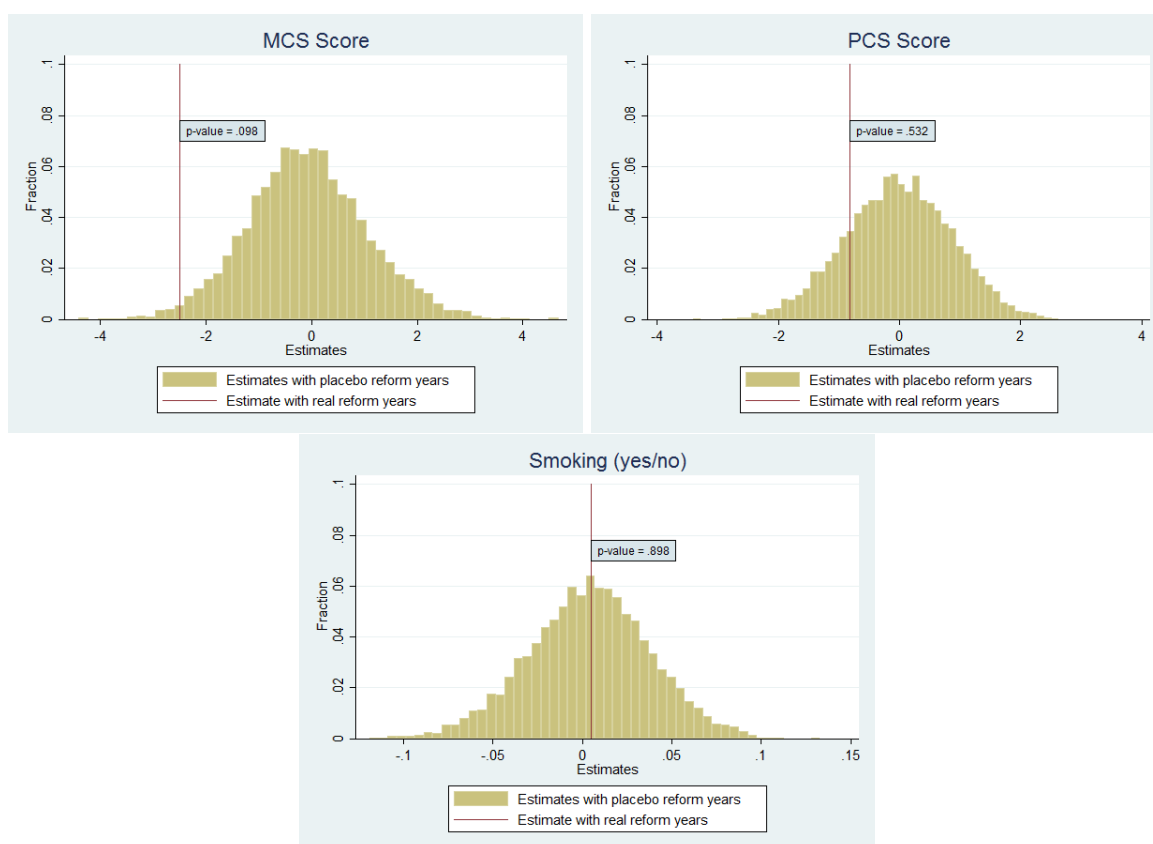
Note: Randomization based on 6,000 permutations of reform dates. Results based on observations from SOEP waves v23-v31 (2006-2014).

7 Discussion

We analyze health effects of a recent school reform in Germany increasing learning intensity in higher secondary schools. Our findings robustly point towards negative reform effects on self-assessed mental health as measured by the aggregate MCS score. According to our main specification, the MCS score of G8 students is about -2.71 points lower than for G9

students. This effect corresponds to about one quarter of a MCS standard deviation. Looking at the literature on specific patient populations, we conclude that this is a sizeable effect. As indicated, for example the cut-off for 30-day episodes of depressive disorders is about 44 % of a standard deviation below the mean 50 (Vilagut et al., 2013). Thus, if we observed a child of a mean MCS score 50 under the old secondary school regime, this same child would move more than halfway towards the depression cut-off when being in a G8 school. We also observe negative physical health effects when considering sub-dimensions of the physical health score (PCS) and in subgroups of our main sample. Additionally, results from our sub-group analyses tentatively suggest that the reform yields higher tobacco consumption for students with less favorable home resources. These students might be more 'vulnerable' students according to the diathesis-stress-model (Ingram and Luxton, 2005). Our general analysis of information related to students' workload and daily activities points to increased (performance) pressure for students under the G8 regime; they more often argue with their parents and are more often attending private tutoring hours.

Our study is limited by the observed period of time since the introduction of the reform. Over 90 % of the students and graduates in our sample belong to the first four cohorts after the implementation of the reform. However, based on the available data, we provide fair evidence on the robustness of our results. Within our period of observation, we find that mental health effects are even more pronounced for more recent cohorts (three or more years after transition to the new regime). Therefore, we conclude that the effects are not solely due to the implementation period. In addition, we observe persistent mental health effects of the reform in young adults after they have graduated from the academic track schools. We thus conclude, that the effects are not limited to temporary shocks in student (mental) health.

Our findings are informative to policy-makers judging on the reform effects and interested in learning from the past. Over the last decade, parents and students claimed that the reform increased stress in schools and exerted detrimental mental health effects on adolescents; we provide evidence to substantiate their assumptions. From a more general point of view, our paper adds insight to the literature on features of the education system affecting human capital development in adolescence. In this regard, the *technology of capability (or skill) formation* (Cunha and Heckman, 2007; Heckman, 2007) suggests that even temporary shocks in skill investments exert lasting effects on the formation of adult human capital.

References

Agüero, J. M. and Beleche, T. (2013). Test-Mex: Estimating the effects of school year length on student performance in Mexico, *Journal of Development Economics* **103**: 353–361.

Almlund, M., Duckworth, A. L., Heckman, J. J. and Kautz, T. D. (2011). Personality psychology and economics, *NBER Working Paper 16822* .

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

Andrietti, V. (2015). The causal effects of increased learning intensity on student achievement: Evidence from a natural experiment, *Universidad Carlos III de Madrid Working Paper Economic Series* (15-06).

Anger, S. and Dahmann, S. (2014). The impact of education on personality: Evidence from a German high school reform, *IZA Discussion Paper* **8139**.

Autorengruppe Bildungsberichterstattung (2008). *Bildung in Deutschland 2008 - Ein indikatorengestützter Bericht mit einer Analyse zu Übergängen im Anschluss an den Sekundarbereich I.*

Autorengruppe Bildungsberichterstattung (2012). *Bildung in Deutschland 2012 - Ein indikatorengestützter Bericht mit einer Analyse zur kulturellen Bildung im Lebenslauf.*

Bellenberg, G. (2012). Schulformwechsel in Deutschland – Durchlässigkeit und Selektion in den 16 Schulsystemen der Bundesländer innerhalb der Sekundarstufe I, *Technical report*, Bertelsmann Stiftung.

Belsky, J. and Pluess, M. (2009). Beyond diathesis stress: Differential susceptibility to environmental influences, *Psychological bulletin* **135**(6): 885–908.

Borghans, L., Duckworth, A. L., Heckman, J. J. and Ter Weel, B. (2008). The economics and psychology of personality traits, *Journal of Human Resources* **43**(4): 972–1059.

Braakmann, N. (2010). A note on the causal link between education and health: Evidence from the German short school years, *University of Lüneburg Working Paper Series in Economics* .

Büttner, B. and Thomsen, S. L. (2013). Are we spending too many years in school? Causal evidence of the impact of shortening secondary school duration, *German Economic Review* **16**(1): 65–86.

Cameron, A. C. and Miller, D. L. (2015). A practitioner's guide to cluster-robust inference, *Journal of Human Resources* **50**(2).

Card, D. and Krueger, A. B. (1992). Does school quality matter? Returns to education and the characteristics of public schools in the United States, *Journal of political Economy* **100**(1): 1–40.

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

Cunha, F. and Heckman, J. J. (2007). The technology of skill formation, *American Economic Review* **97**(2): 31–47.

Cutler, D. M. and Lleras-Muney, A. (2012). Education and health: Insights from international comparisons, *NBER Working Paper 17738* .

Dörsam, M. and Lauber, V. (2015). The effect of a compressed high school curriculum on university performance, mimeo.

Engels, D. and Thielebein, C. (2011). *Zusammenhang von sozialer Schicht und Teilnahme an Kultur-, Bildungs- und Freizeitangeboten für Kinder und Jugendliche*, ISG Institut für Sozialforschung und Gesellschaftspolitik.

Feinstein, L., Sabates, R., Anderson, T. M., Sorhaindo, A. and Hammond, C. (2006). What are the effects of education on health? Measuring the effects of education on health and civic engagement, *Proceedings of the Copenhagen Symposium, OECD*.

Firpo, S., Fortin, N. M. and Lemieux, T. (2009). Unconditional quantile regressions, *Econometrica* **77**(3): 953–973.

Fischer, M., Karlsson, M., Nilsson, T. and Schwarz, N. (2016). The sooner the better? compulsory schooling reforms in Sweden, *IZA Discussion Paper No. 10430* .

Gill, S. C., Butterworh, P., Rodgers, B. and Mackinnon, A. (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**: 63 –71.

Glemser, A., Huber, S. and Bohlender, A. (2014). *SOEP 2014 – Methodenbericht zum Befragungsjahr 2014 (Welle 31) des Sozio-oekonomischen Panels*, DIW Berlin.

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

Grossman, M. (2015). The relationship between health and schooling: What's new?, *NBER Working Paper 21609* .

Heckman, J. J. (2007). The economics, technology, and neuroscience of human capability formation, *Proceedings of the National Academy of Sciences of the United States of America* **104**(33): 13250–13255.

Hille, A., Spieß, K. and Staneva, M. (2016). *Immer mehr Schülerinnen und Schüler nehmen Nachhilfe, besonders in Haushalten mit mittlerem Einkommen*, DIW Berlin.

Homuth, C. (2017). *Die G8-Reform in Deutschland*, PhD thesis, Otto-Friedrich-Universität Bamberg.

Huebener, M., Kuger, S. and Marcus, J. (2017). Increased instruction hours and the widening gap in student performance, *Labour Economics* **47**: 15–34.

Huebener, M. and Marcus, J. (2015a). Empirische Befunde zu Auswirkungen der G8-Schulzeitverkürzung, *Diw Roundup 57*, DIW Berlin.

Huebener, M. and Marcus, J. (2015b). Moving up a gear: The impact of compressing instruction time into fewer years of schooling, *DIW Discussion Paper 1450* .

Ingram, R. E. and Luxton, D. D. (2005). Vulnerability-stress models, *in* B. L. Hankin and R. Abela, John (eds), *Development of psychopathology: A vulnerability-stress perspective*, Sage Publications, pp. 32–46.

Jacobshagen, N. and Rigotti, T. (2008). Voll der Stress? Stress bei Adoleszenten in den letzten Schuljahren, *in* A. Krause, H. Schüpback, E. Ulich and M. Wülser (eds), *Arbeitsort Schule. Organisations- und arbeitspsychologische Perspektiven*, Gabler, chapter 11, pp. 289–308.

Konsortium Bildungsberichterstattung (2006). *Bildung in Deutschland - Ein indikatorengestützter Bericht mit einer Analyse zu Bildung und Migration*.

Krashinsky, H. (2014). How would one extra year of high school affect academic performance in university? Evidence from an educational poliy change, *Canadian Journal of Economics* **47**(1): 70–97.

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

Lavy, V. (2016). Expanding school resources and increasing time on task: Effects on students' academic and non-cognitive outcomes, mimeo.

Mannan, M., Mamun, A., Doi, S. and Clavarino, A. (2016). Prospective associations between depression and obesity for adolescent males and females – a systematic review and meta-analysis of longitudinal studies, *PLoS ONE* **11**(6).

Meyer, T. and Thomsen, S. L. (2015). Eine Evaluation der Wirkungen der verkürzten Gymnasialschulzeit auf die außerschulischen Aktivitäten der Schülerinnen und Schüler, *Schmollers Jahrbuch* **135**: 249 – 278.

Meyer, T. and Thomsen, S. L. (2016). How important is secondary school duration for postsecondary education decisions? Evidence from a natural experiment, *Journal of Human Capital* **10**(1): 67–108.

Meyer, T., Thomsen, S. L. and Schneider, H. (2016). New evidence on the effects of the shortened school duration in the German states: An evaluation of postsecondary education decisions, mimeo.

Milde-Busch, A., Blaschek, A., Borggräfe, I., von Kries, R., Straube, A. and Heinen, F. (2010). Besteht ein Zusammenhang zwischen der verkürzten Gymnasialzeit und Kopfschmerzen und gesundheitlichen Belastungen bei Schülern im Jugendalter?, *Klinische Pädiatrie* **222**: 1–6.

Minkley, N., Rest, M., Terstegen, S., Kirchner, W. H. and Wolf, O. T. (2015). Mehr Stress durch G8? Stressbelastung von Abiturienten mit regulärer und verkürzter Gymnasialzeit in NRW, *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie* **47**(4): 188–197.

Mühlenweg, A. M. and Puhani, P. A. (2010). The evolution of the school-entry age effect in a school tracking system, *Journal of Human Resources* **45**(2): 407–438.

Müller-Nordhorn, J., Roll, S. and Willich, S. N. (2004). Comparison of the Short Form (SF)-12 health status instrument with the SF-36 in patients with coronary heart disease, *Heart* **90**(5): 523–527.

OECD (2002). *Education at a glance 2002*, Organization for Economic Co-operation and Development.

Patall, E. A., Cooper, H. and Batts Allen, A. (2010). Extending the school day or school year: A systematic review of research (1985-2009), *Review of Educational Research* **80**(3): 401–436.

Pischke, J.-S. (2007). The impact of length of the school year on student performance and earnings: Evidence from the German short school years, *The Economic Journal* **117**: 1216–1242.

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

Quis, J. and Reif, S. (2017). Health effects of instruction intensity – evidence from a natural experiment in German high-schools, *FAU Discussion Paper No.12/2017* .

Ramseier, E., Allraum, J., Stalder, U., Grin, F., Alliata, R., Müller, S., Willimann, I., Dozio, E., Labudde, P., Maag Merki, K. et al. (2004). *Evaluation der Maturitätsreform 1995 (EVA-MAR). Neue Fächerstruktur, pädagogische Ziele, Schulentwicklung – Schlussbericht zur Phase I*, Schweizerische Konferenz der kantonalen Erziehungsdirektoren (EDK); Bundesamt für Bildung und Wissenschaft (BBW).

Schupp, J. and Wagner, G. G. (2007). *New Concepts and Survey Questions in the German Socio-Economic Panel Study (SOEP) - Principles*, DIW Berlin Data Documentation.

Shankar, N. L. and Park, C. L. (2016). Effects of stress on students' physical and mental health and academic success, *International Journal of School & Educational Psychology* **4**(1).

Thiel, H., Thomsen, S. L. and Büttner, B. (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): 861–892.

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

Trautwein, U., Hübner, N., Wagner, W. and Kramer, J. (2015). *Konsequenzen der G8-Reform. Eine Studie über Leistungen, Wohlbefinden und Freizeitverhalten von Schülerinnen und Schülern vor und nach der G8-Reform in Baden-Württemberg*, Eberhard Karls Universität Tübingen.

Vilagut, G., Forero, C. G., Pinto-Meza, A., Haro, J. M., De Graaf, R., Bruffaerts, R., Kovess, V., De Girolamo, G., Matschinger, H., Ferrer, M. et al. (2013). The mental component of the Short-Form health Survey (SF-12) as a measure of depressive disorders in the general population: results with three alternative scoring methods, *Value in Health* **16**(4): 564–573.

Weidner, G., Kohlmann, C.-W., Dotzauer, E. and Burns, L. R. (1996). The effects of academic stress on health behaviors in young adults, *Anxiety, Stress, & Coping* **9**(2): 123–133.

Westermaier, F. G. (2016). The impact of lengthening the school day on substance abuse and crime: Evidence from a German high school reform, *DIW Discussion Paper 1616* .

WHO (2017). *Global Accelerated Action for the Health of Adolescents (AA-HA!) – Guidance to Support Country Implementation*, World Health Organization.

Wößmann, L., Lergetporer, P., Kugler, F. and Werner, K. (2014). Was die Deutschen über die Bildungspolitik denken – Ergebnisse des ersten ifo Bildungsbarometers, *Ifo Schnelldienst* **67**(18): 16–33.

Ziebarth, N. (2010). Measurement of health, health inequality, and reporting heterogeneity, *Social Science & Medicine* **71**(1): 116–124.

Appendix

Table A-1: Description of background variables

Variable	Description
Age	Age in years
Male	Male
Migration background	Either the student or at least one parent not born in Germany
High parental education	At least one parent with higher secondary education
Non-working mother	Mother not working (while student living at home)
Single child	Student is a single child
First born child	Student is the first born child
Single parent	Student did not grow up with both parents until the age of 15
Rural	Student grew up in the countryside until the age of 15

Table A-2: Balancing tests on background variables

| | Mean | Correlative | | | Diff.-in-Diff. | |
		(I)	(Ia)	(Ib)	(II)	(III)
Male	0.47	-0.044	-0.046 *	-0.002	-0.002	-0.001
		(0.027)	(0.027)	(0.015)	(0.014)	(0.014)
Migration background	0.20	0.086 **	0.140 ***	-0.040 **	0.010	0.008
		(0.034)	(0.033)	(0.019)	(0.017)	(0.018)
Non-working mother	0.12	0.107 **	0.126 ***	0.015	0.032	0.032
		(0.042)	(0.041)	(0.024)	(0.021)	(0.022)
Single child	0.11	-0.024	-0.034	0.006	-0.017	-0.016
		(0.044)	(0.043)	(0.025)	(0.022)	(0.025)
First born child	0.58	0.093 ***	0.091 ***	0.019	0.008	0.015
		(0.028)	(0.027)	(0.016)	(0.014)	(0.015)
Single parent	0.22	0.101 ***	0.100 ***	0.008	0.003	0.004
		(0.033)	(0.032)	(0.019)	(0.017)	(0.017)
Rural	0.24	0.093 ***	0.039	0.047 ***	-0.011	-0.012
		(0.032)	(0.032)	(0.018)	(0.017)	(0.017)
Fed. state indicators		no	yes	no	yes	yes
Cohort indicators		no	no	yes	yes	yes
Control variables						
– Age		yes	yes	yes	yes	yes
– Backgr. characteristics		no	no	no	no	yes
N		1, 130	1, 130	1, 130	1, 130	1, 130

Heading row above the sub-columns: **Outcome: Treatment (G8 participation)** spanning Correlative and Diff.-in-Diff.

Note: Balancing tests on background variables. Estimates based on observations from SOEP waves v23-v31 (2006-2014). Specifications: (I) Baseline specification: OLS regression of treatment (G8 reform) on background variables (separate regressions) including age and age squared as controls; (Ia) Baseline specification including federal state indicators; (Ib) Baseline specification including school entry cohort indicators; (II) Baseline specification including both federal state *and* school entry cohort indicators; (III) Baseline specification including federal state and school entry cohor indicators and including the full set of background variables as controls. Cluster-robust standard errros in parentheses. * $p < 0.1$; ** $p < 0.05$; ** $p < 0.01$. The balancing tests do not point to any relevant differences between G9 and G8 cohorts after controlling for age, federal state and school entry cohort (specification II). Once these are controlled for, including further background variables as controls does not change the coefficients (specification III).

Table A-3: Age distribution and federal state of residence of studends in G9 and G8 scheme

|	G9 students	G8 students
Age group		
18-19	438	357
20-21	152	54
>22	125	4
Federal state		
Saxony-Anhalt	29	36
Mecklenburg West.-Pom.	21	25
Hamburg	10	15
Bavaria	111	101
Baden-Württemberg	141	64
Berlin	46	20
Lower Saxony	91	72
North Rhine-Westphalia	266	82

Note: Based on 1,130 observations from SOEP waves v23-v31 (2006-2014).

Table A-4: Interactions between G8 reform and background variables

	MCS Score	PCS Score	Smoking (yes/no)	Smoking (>5 cig. per day)	Smoking (>15 cig. per day)
G8 effect	-3.231	-0.798	-0.009	0.039	0.041
	(2.107)	(1.074)	(0.068)	(0.053)	(0.035)
Interactions terms (G8 effect # background characteristic)					
Male	0.390	-0.549	-0.039	-0.041	-0.021
	(1.065)	(1.163)	(0.030)	(0.039)	(0.029)
Migration background	-0.758	-0.665	0.066*	0.077**	0.049
	(1.299)	(1.288)	(0.031)	(0.032)	(0.035)
High parental educ.	-1.444	0.295	-0.007	-0.020	-0.044**
	(1.757)	(0.658)	(0.046)	(0.036)	(0.018)
Non-working mother	-1.077	1.974	0.016	-0.023	0.061***
	(2.613)	(1.285)	(0.072)	(0.021)	(0.014)
Single child	-4.468*	0.036	-0.032	0.016	-0.009
	(2.310)	(1.429)	(0.138)	(0.113)	(0.044)
First born child	3.113**	-0.862	0.030	-0.001	-0.008
	(1.186)	(0.539)	(0.035)	(0.026)	(0.017)
Single parent	0.428	0.576	0.009	0.017	0.017
	(1.298)	(0.703)	(0.041)	(0.047)	(0.024)
Rural	-0.751	0.901	0.026	-0.001	-0.008
	(0.666)	(1.005)	(0.043)	(0.044)	(0.016)
Fed. state indicators	yes	yes	yes	yes	yes
Cohort indicators	yes	yes	yes	yes	yes
Control variables					
– Age	yes	yes	yes	yes	yes
– Backgr. characteristics	yes	yes	yes	yes	yes
N	$1,117$	$1,117$	$1,105$	$1,105$	$1,105$

Note: Results based on observations from SOEP waves v23-v31 (2006-2014). OLS regressions of health outcome on treatment (G8 participation) including interaction terms between each control variable and the G8-indicator and including both federal state and school entry cohort indicators as well as the full set of background variables as controls. Cluster-robust standard errors in parentheses, * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

Figure A-1: Education system in Germany (stylized illustration, before G8 reform)
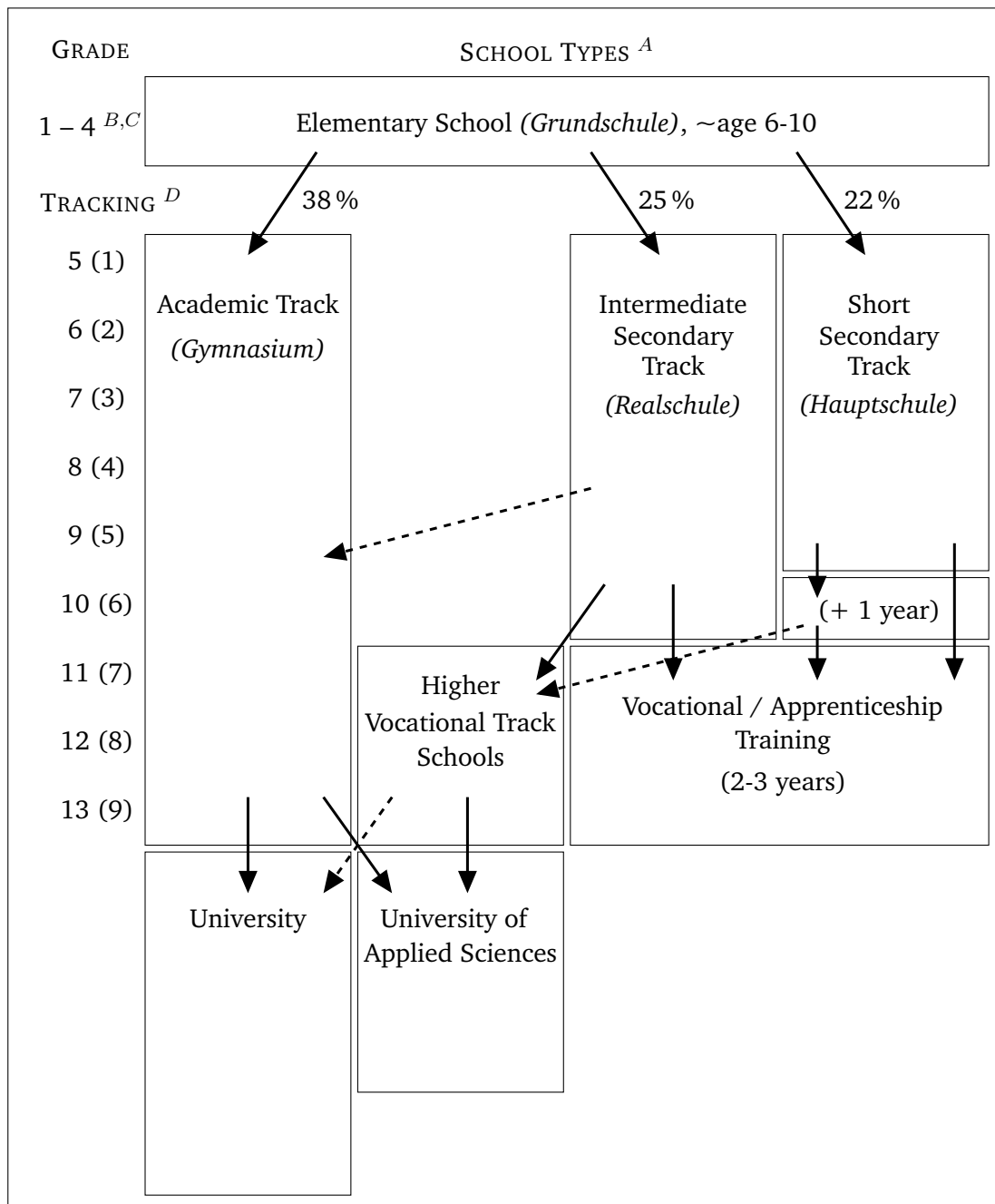
Note: [A] Education system before G8 reform. [B] In orientation stages, tracking is delayed to grade 6. [C] Numbers in left panel relate to grades, numbers in parenthesis to secondary education grades. [D] Percentages are percentages of students tracked to respective school types in the median/mode year of reform implementation (2004). For simplicity, comprehensive and combined secondary schools (8 % and 6 % of students) and orientation stage students (2 %) are not included. *Source:* Own presentation, based on Autorengruppe Bildungsberichterstattung (2012), page XI and stylized figure in Mühlenweg and Puhani (2010). Proportion of students in school types according to Konsortium Bildungsberichterstattung (2006).

Table A-5: Questions in the SOEP questionnaire and sub-scales behind the MCS and PCS scales

Question in SOEP	Subscale
Mental Component Summary Scale (MCS)	
• During the last 4 weeks, how often did you feel energetic?	→ Vitality
• During the last 4 weeks, how often did you feel that due to physical or mental health problems you were limited that is, in contact with friends, acquaintances, or relatives?	→ Social Interaction
• During the last 4 weeks, how often did you feel that due to mental health or emotional problems you achieved less than you wanted to at work or in everyday activities?	⎫
• During the last 4 weeks, how often did you feel that due to mental health or emotional problems, you carried out your work or everyday tasks less thoroughly than usual?	⎬ Work ability (mental)
• During the last 4 weeks, how often did you feel down and gloomy?	⎫
• During the last 4 weeks, how often did you feel calm and relaxed?	⎬ Emotional Balance
Physical Component Summary Scale (PCS)	
• How would you describe your current health?	→ General health
• When you have to climb several flights of stairs on foot, does your health limit you greatly, somewhat, or not at all?	⎫
• And what about other demanding everyday activities, such as when you have to lift something heavy or do something requiring physical mobility: Does your health limit you greatly, somewhat, or not at all?	⎬ Functional ability
• During the last 4 weeks, how often did you feel that due to physicall health problems you achieved less than you wanted to at work or in everyday activities?	⎫
• During the last 4 weeks, how often did you feel that due to physical health problems, you carried out your work or everyday tasks less thoroughly than usual?	⎬ Work ability (physical)
• During the last 4 weeks, how often did you have severe physical pain?	→ Bodypain

Source: SOEP 2014 Living in Germany – Survey 2014 on the social situation of households (*Erhebungsinstrumente 2014 (Welle 31) des Sozio-oekonomischen Panels: Personenfragebogen, Altstichproben*).