

Marcus, Jan; Reif, Simon; Wuppermann, Amélie; Rouche, Amélie

**Article — Accepted Manuscript (Postprint)**

## Increased instruction time and stress-related health problems among school children

Journal of Health Economics

**Provided in Cooperation with:**

German Institute for Economic Research (DIW Berlin)

*Suggested Citation:* Marcus, Jan; Reif, Simon; Wuppermann, Amélie; Rouche, Amélie (2020) : Increased instruction time and stress-related health problems among school children, Journal of Health Economics, ISSN 0167-6296, Elsevier BV, Amsterdam, Vol. 70, <https://doi.org/10.1016/j.jhealeco.2019.102256>, <https://www.sciencedirect.com/science/article/abs/pii/S0167629619303467?via%3Dihub>

This Version is available at:

<https://hdl.handle.net/10419/250057>

**Standard-Nutzungsbedingungen:**

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

**Terms of use:**

*Documents in EconStor may be saved and copied for your personal and scholarly purposes.*

*You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.*

*If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.*



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

# Increased instruction time and stress-related health problems among school children

Jan Marcus\*, Simon Reif†, Amelie Wuppermann‡, Amélie Rouche§

September 2019

## Abstract

While several studies suggest that stress-related mental health problems among school children are related to specific elements of schooling, empirical evidence on this causal relationship is scarce. We examine a German schooling reform that increased weekly instruction time and study its effects on stress-related outpatient diagnoses from the universe of health claims data of the German Social Health Insurance. Exploiting the differential timing in the reform implementation across states, we show that the reform slightly increased stress-related health problems among school children. While increasing instruction time might increase student performance, it might have adverse effects in terms of additional stress.

*Keywords:* stress, mental health, instruction time, G8 reform

*JEL:* I18, I28

---

\*University of Hamburg & DIW Berlin ([jan.marcus@uni-hamburg.de](mailto:jan.marcus@uni-hamburg.de))

†University of Erlangen-Nuremberg & RWI Essen ([simon.reif@fau.de](mailto:simon.reif@fau.de))

‡MLU Halle-Wittenberg ([amelie.wuppermann@wiwi.uni-halle.de](mailto:amelie.wuppermann@wiwi.uni-halle.de))

§Charité - University Hospital Berlin ([amelierouche@gmail.com](mailto:amelierouche@gmail.com))

The data used in this study are provided by the Central Research Institute of Ambulatory Health Care in Germany (*Zentralinstitut für die kassenärztliche Versorgung*; Zi). We are particularly grateful to Jörg Bätzing of the Zi for his valuable support with the data. We also want to thank Mirjam Reutter for valuable comments and Adam Lederer for language editing. Funding through the International Doctoral Program, “Evidence-Based Economics,” of the Elite Network of Bavaria is gratefully acknowledged.

# 1 Introduction

Child wellbeing is long on the agenda of both researchers and policy makers (WHO, 2016). Mental health problems, in general, and stress, in particular, are identified as detrimental factors to wellbeing in childhood and adolescents (Bor et al., 2014) with possible negative consequences for later life outcomes (Hughes et al., 2017; Felitti et al., 2019). At the same time mental health problems are increasing among children and adolescents (Collishaw et al., 2004; Collishaw, 2015).

Several studies argue that schooling substantially contributes to the high stress prevalence and stress-related mental health problems. For example, youth suicide rates increase at the beginning of the school year (Lahti et al., 2005; Matsubayashi et al., 2016) and, generally, are higher while school is in session (Hansen and Lang, 2011). Further, there is evidence that academic stress increases suicidal ideation (Ang and Huan, 2006). Although stress in school rarely leads to extreme outcomes, such as (attempted) suicide, students report that school is the biggest stressor in their lives (Elias, 1989; de Anda et al., 2000; Baltimore Center for Adolescent Health, 2006). However, it is unclear which elements of schooling and the schooling system drive the observed patterns.<sup>1</sup>

In this paper, we contribute to the literature on schooling and mental health by investigating the causal effect on stress-related mental health problems of a specific element of the schooling system, namely weekly instruction time. More specifically, we exploit a German schooling reform that substantially increased weekly instruction time. The differential timing of the reform across German states allows us to estimate the causal effect of more weekly instruction time using a difference-in-differences approach. Based on the universe of outpatient diagnoses from health claims data of the German Social Health Insurance system, we find evidence that more instruction time increases stress-related outpatient diagnoses slightly, albeit statistically significantly.

Several empirical studies document associations between schooling and students' self-reported mental health.<sup>2</sup> For instance, Leung et al. (2010) report that high academic pressure is associated with high anxiety levels. Similarly, Löfstedt (2017) finds that increased pressure in school relates to higher rates of psychosomatic symptoms. Freeman et al. (2012) find that a negative school climate correlates with high levels

---

<sup>1</sup>Apart from elements of the schooling system, peers might also affect mental health and stress of students, e.g., through bullying (Eriksen et al., 2014). Further, schooling might not only have negative consequences for stress and mental health, but it might also have protecting effects. For instance, children who experience a lot of stress at home might be more stressed during school holidays, with school potentially being a place to gain resilience.

<sup>2</sup>All these studies focus on the effect of schooling on mental health when students are in school. There is also a distinct literature focusing on the long-run effects of education on mental health (as a potential non-monetary return to education). These studies have a different focus compared to our study, as we are mainly interested in explaining youth mental health issues.

of psychosomatic complaints. Further, Torsheim and Wold (2001) show that somatic symptoms, such as headache, abdominal pain, backache, and dizziness, are common when perceived school-related stress is high. In line with these results are the findings of Ravens-Sieberer et al. (2009), who show that there is a positive relationship between better school perceptions and higher self-assessed health. Other studies focus on associations between specific elements of the schooling system and measures of mental health. For instance, Galloway et al. (2013) document that students are more stressed when they need more time to complete their homework. Similarly, there is evidence that students are more stressed during exam periods (Banks and Smyth, 2015), while the amount of school sport (Jewett et al., 2014) and good teacher-student relationships (Conner et al., 2014) relate to better self-reported measures of mental health.

There is also research on the causal effects of schooling on measures of mental health. For instance, Song (2017) applies a regression discontinuity design in the admission to elite colleges in China and finds that attending an elite college increases self-reported measures of stress. Heissel et al. (2018) provide evidence that cortisol levels, as an indicator of stress, increase during high-stakes testing weeks. King et al. (2014) demonstrate that adolescents in the US receive more stimulant prescriptions during the school year. They further show that states with stricter accountability laws have the largest differences in stimulant prescriptions between school year and summer, which suggests that there is a high level of self-medication in states with less restrictive accountability laws. Results from Bokhari and Schneider (2011) support this claim by showing that stricter accountability laws increase diagnosis and medication for ADHD.<sup>3</sup> The studies by Bokhari and Schneider (2011) and King et al. (2014) are most similar to our study regarding the aim to identify causal effects and the focus on more objective health measures that are not self-reported by the individuals. However, while these studies examine differences between school and holiday season and effects of school accountability laws, we focus on the effects of a different element of the schooling system, instruction time. Moreover, these studies look at psychostimulant prescription, we study stress-related diagnoses.

This paper contributes to the existing literature along three dimensions. First, while many previous studies focus on correlations, our difference-in-differences strategy aims to identify causal effects. Second, instead of examining self-reported stress-related mental health problems, we rely on outpatient diagnoses of stress-related diseases and, thus, a more objective measure. Third, we do not look at schooling in general but rather focus on one key feature of the schooling system – weekly instruction time. The remainder

---

<sup>3</sup>Additionally, students might be prescribed drugs due to unobservable factors, such as their relative age, independent of stress in school. Schwandt and Wuppermann (2016), for example, show that ADHD medication is more frequently prescribed when children start school younger.

of this paper is structured as follows: Next we provide an overview of the schooling reform we exploit (Section 2). Section 3 outlines the data we use and the empirical strategy, while Section 4 presents the results and robustness checks. Section 5 concludes.

## 2 Institutional Background

In order to analyze the mental health effects of weekly instruction time, we exploit a major reform in German secondary schools. In the course of this reform, the majority of German states shortened their academic track from nine to eight years, reducing total school years from 13 to 12 (the so called G8 reform). This reduction in school years took place early in the new millennium and was compensated by an increase of weekly instruction time in the remaining school years. For grades 5-10, the focus of this paper, this increase amounts to one to three schooling hours (each 45 minutes) per week (see Figure 1). Across grades 5-10, weekly instruction time increased by slightly more than two schooling hours on average, which amounts to less than 10 % of the overall weekly instruction time, which is about 30 hours. The increase in weekly instruction time is used to compensate for the one year reduction in school years by spreading content over earlier school years. Naturally, the G8 reform increased the speed at which students must cope with the content learned as some material from higher grades was now covered in earlier grades. Hence, the reform entails both an increase in weekly instruction time and additional learning content, and any reform effects that we estimate capture effects of increases in instruction time that is filled with additional learning content rather than just additional hours that children spent at school.

The switch from the previous regime (G9) to the new regime (G8) affected only the academic track school (*Gymnasium*), which is the high ability track that prepares for university studies. This track is attended by about 35 % of a cohort according to representative survey data (see Table A.1 in the Appendix). Weekly instruction time and the length of the schooling period remained constant in the other school tracks.

Education policy is in the domain of the sixteen German states. The G8 reform constitutes a natural experiment as different federal states implemented it in different years (see Table 1). The eastern German states of Saxony-Anhalt and Mecklenburg-Vorpommern were the first to adopt the G8 system. Both states had a G8 system in the 1990s and before reunification. The next state to implement the reform was Saarland. In Saarland, with its strong ties to neighboring France, policymakers were concerned that their graduates were at a disadvantage compared to the French graduates, who were younger at graduation. Five states implemented the G8 reform with the first G8 graduates in 2012. As by construction of the G8 reform the last G9 cohort and the first

G8 cohort graduate at the same time (this is the so-called double graduation cohort), it makes sense that North Rhine-Westphalia, Germany's most populous state, decided to "produce" its first G8 graduates (and, therefore, the double graduation cohort) one year later. Overall, the decisions to switch to the G8 system were not based on health- or stress-related considerations.

This policy experiment is already exploited in several studies, mainly regarding students' behavior and academic outcomes. For instance, Huebener et al. (2017) find that the reform increased academic performance in school due to the higher weekly instruction time in the lower grades. However, this effect is mainly driven by high-performing students; lower-performing students benefit less from the additional instruction time. The finding that some students have trouble in coping with the additional instruction time and learning material is underlined by the finding of Huebener and Marcus (2017), who show that more students repeat a grade due to the reform. This finding also can be interpreted as suggestive evidence that the G8 reform increases stress. At the end of schooling, several studies document slightly lower performance of the G8 students (Büttner and Thomsen, 2015; Huebener and Marcus, 2017; Hübner et al., 2017), suggesting that the additional instruction time in lower grades does not fully compensate for the omitted last year. Further, there is some evidence that the reform affected cognitive skills (Dahmann, 2017), but only marginally affected personality traits (Dahmann and Anger, 2014; Thiel et al., 2014). Affected students had also less leisure time (Meyer and Thomsen, 2015; Hübner et al., 2017). Further, the reform resulted in delayed university enrollment for some students (Meyer and Thomsen, 2016; Marcus and Zambre, 2019) and slightly reduced enrollment rates (Marcus and Zambre, 2019). However, affected students graduated earlier from high school (Huebener and Marcus, 2017) and enrolled in university at significantly younger ages (Marcus and Zambre, 2019), thereby allowing for an earlier labor market entry.

While existing research focuses mainly on students' behavior and academic outcomes, parents and students were strongly concerned about higher stress levels due to the reform's increase in weekly instruction time. These concerns – although based on subjective feelings and anecdotal evidence – were articulated widely in the media (see, e.g., F.A.Z., 2014; Die Welt, 2016; Spiegel Online, 2016) and prompted several states to repeal the reform (see Table 1). Other states are also considering repealing the reform. The decisions to switch back were not based on scientific studies (Huebener and Marcus, 2015; Thomsen and Anger, 2018), although there is initial evidence on the health effects of the reform based on survey data.

Milde-Busch et al. (2010) are the first to analyze the health effects of the G8 introduction. They conducted a survey in different schools when both G8 and G9 students

were still in school and find no evidence for health differences. A similar empirical design is used by Quis (2018) with large scale survey data from the National Educational Panel Study (NEPS), where she compares students in G8 and G9 at the end of their time in school and finds some evidence for more mental health problems in G8. These studies do not control for age effects since the data was only collected at one point in time. Quis and Reif (2017) and Hofmann and Mühlenweg (2018) use survey data from the Socio-Economic Panel (SOEP) for 17 year old students in a difference-in-differences framework similar to the one in this paper. The results indicate small negative effects of G8 on subjective measures of mental health. However, these subjective measures of students' health could be influenced by parental opinions and the general public's ongoing discussion regarding the reform.<sup>4</sup> To date, no empirical study analyzes the effects of increased instruction intensity on objective stress-related health measures. In this paper, we extend the existing survey evidence by estimating the effect of increased instruction time on stress-related health problems of students by using administrative data.

### 3 Data and empirical strategy

For our outcome measure, we make use of administrative data on outpatient diagnoses. The data is provided by *Zentralinstitut für die kassenärztliche Versorgung* (Zi), on behalf of the National Association of Statutory Health Insurance Physicians, and covers the universe of outpatient diagnoses from health claims data of the German Social Health Insurance (SHI). The German SHI covers about 70 million individuals (German Federal Ministry of Health, 2018), which amounts to about 87 % of the population across all age groups.<sup>5</sup> Our data does not include information about doctor visits by privately insured individuals. Further, doctor visits due to accidents at school/work or on the way to school/work are not covered by the SHI system and, therefore, not included in our data set.<sup>6</sup> Our data includes information about an individual's outpatient diagnoses (ICD-10 codes) in a given calendar year, the individual's month and year of birth, as well as the county and state of residence. We cannot link individuals across years as the data does not include an individual panel identifier due to data confidentiality. Hence, our data can be rather seen as pooled cross-sectional data. The applied difference-in-differences

---

<sup>4</sup>There is plenty of anecdotal evidence that children affected by the reform were repeatedly asked by parents and relatives whether they feel stressed by this new system, which was widely discussed in the media.

<sup>5</sup>Survey data from the KiGGS (RKI, 2015) indicates that the share of children aged 18 or younger who are insured in the SHI is similar, at about 89 % (see Table A.1 in the Appendix).

<sup>6</sup>These doctor visits are covered by the mandatory accident insurance.

identification strategy does not require panel data.<sup>7</sup>

The data set has two major strengths. The first is its large sample size: Our main analyses are based on more than 2.7 million children per year. The second advantage is that the relevant health information are provided by physicians.<sup>8</sup> This administrative, objective, stress measure complements previous studies that use subjective health indicators (Quis and Reif, 2017; Hofmann and Mühlenweg, 2018; Quis, 2018). While subjective measures have the advantage of capturing also small changes in health, our objective measure indicates strong, medically relevant stress levels and, therefore, is less likely to be prone to reporting error, social desirability bias, or mood effects.

The data also has limitations. Data is only available for a limited number of years (2008–2011). Further, the data set does not provide any information on school track and school grade. While the G8 reform affected only the academic track, we have to assign the treatment indicator irrespective of the school track attended. Therefore, our estimates are likely biased toward zero and should be interpreted as intent-to-treat effects. In the robustness section, we discuss this issue in more detail and provide estimates for the effects if only academic-track students were considered. The information on the school grade is important for correctly assigning the G8 treatment indicator. We assign the school grade (and, hence, the treatment indicator) based on the individual's date of birth. This assignment rule assumes that individuals complied with the school entry rules and did not repeat or skip a grade. For cohorts directly before and after the introduction of the G8 reform, both non-compliance with the entry rules and grade repetition/skipping can lead to a misassignment of the reform indicator and, hence, a bias toward zero. In Section 4.2, we show that the results are robust to excluding the first G8 or the last G9 cohort. Generally, the extent of the misassignment is clearly smaller than the misassignment resulting from the unknown track information.<sup>9</sup>

For additional computations, we use (i) survey data from the first wave of KiGGS, a nationwide health-survey among children conducted by the Robert Koch Institute in the years 2003-2006 (RKI, 2015).<sup>10</sup>, (ii) state-level information provided by the German Federal Statistical Office on the number of students in different tracks (German Federal

---

<sup>7</sup>In the absence of non-response (as in the present case), the difference-in-differences (DiD) and the fixed effects estimator are asymptotically equivalent. There is a discussion whether and under which circumstances a DiD estimator with individual fixed effects is preferred to a DiD estimator without individual fixed effects when there is non-response (see Lechner et al., 2016).

<sup>8</sup>In Germany, physicians in outpatient care do not have financial incentives to assign specific diagnoses. Unlike in the German hospital sector, reimbursement does not depend on the diagnoses.

<sup>9</sup>Compliance with the age cut-offs for school entry is, at 86 % (see Schwandt and Wuppermann, 2016), rather high in Germany. Further, the average grade repetition rate in academic-track schools is less than 2 % in grades 5-10 (see German Federal Statistical Office, 2012). Huebener and Marcus (2017) also provide evidence that the G8 reform did not increase grade repetition before grade 10.

<sup>10</sup>We use the KiGGS survey only for illustrative purposes as the number of observation at the state-grade level is too small for an analysis of the reform effect.



Statistical Office, 2012) and on the annual expenses per student<sup>11</sup>, and (iii) county-level information on the numbers of psychologists, pediatricians, and general practitioners provided by the German Federal Institute for Research on Building, Urban Affairs and Spatial Development (BBSR).

### 3.1 Sample

We restrict our main analysis sample to individuals in grades 7 to 10. This restriction is imposed for several reasons. First, some states track students only after six grades of joint primary schooling. Hence, by restricting our main sample to individuals in grades 7 and higher, we make sure that we look at an age range in which tracking has taken place in all states. Second, Figure 1 shows that the increase in instruction time is much greater in grades 7 to 10 than in grades 5 and 6. Third, schooling in Germany is compulsory through grade 10 in most states (Vossenkuhl, 2010) and in academic track schools almost all students stay at least until grade 10. After grade 10, some students decide not to pursue the general university entrance qualification but instead complete an apprenticeship. Hence, if the G8 reform impacts the share of academic track students who decide not to pursue the general university entrance qualification, students in grades 11 and 12 might be a selective group. Fourth, in the final two years at academic track schools students earn grades that count toward their final grade point average. This final grade point average is very important for the students because it is a major selection criterion for university admission. Therefore, the last two years might be especially stressful for students. However, under the G8 regime, grades 11 and 12 constitute the final two years, while the final years under the old system were grades 12 and 13. Hence, any differences in stress-related diagnoses in a comparison of students in grade 11 under G8 and the old system might not result from additional instruction time but rather from the fact that the grades already count toward the final grade point average. Therefore, we do not consider students in grades 11 and 12 in our main specification. In the robustness section, we show, however, that our results are insensitive to including students from further grades as well. We also exclude individuals from the state of Hesse in our main analysis as this state gradually implemented the reform across three cohorts (see Table 1). However, when we include Hesse our conclusions do not change (see Section 4.2). Overall, our main sample consists of more than 2.7 million children per year in grades 7-10.

---

<sup>11</sup>This variable is based on information provided by the German Federal Statistical Office (2014) for the *Sekundarstufe* (i.e. up to grade 10), weighted by size of school type in the respective state.

### 3.2 Outcome Measure

Identifying stress-related health problems is challenging as stress is related to different health problems (Torsheim and Wold, 2001; Ravens-Sieberer et al., 2009). In our analysis, we use an index that captures and aggregates different possible health problems related to stress: The Disease based Stress Score (DbSS). The Stress Score, developed by Amélie Rouche at Zi (Rouche et al., 2017b), is based on outpatient diagnoses by physicians. It does not measure stress directly, instead it uses diagnoses that, according to clinicians, indicate physiological stress. As such, it is a general purpose measure that is not geared specifically toward any age bracket. While the relationship between stress and the selected diagnoses is generally quite stable throughout the lifespan, some diagnoses are less prevalent among children (e.g., alcohol abuse). The Stress Score is constructed in four steps. In the first step, stress-related diagnoses (i.e., ICD-10 codes) are identified by experts from medicine, psychology, and epidemiology. The second step groups these more than 300 ICD-10 codes into 21 categories. The third step constructs, for each category, a binary variable that indicates whether an individual had at least one diagnosis in this category in the given calendar year. We refer to these 21 indicator variables as subscales.<sup>12</sup> In supplementary empirical analyses, we also use these indicator variables as outcome variables.<sup>13</sup> The fourth step constructs the actual Stress Score by summing over all these indicator variables. Hence, the range of possible values of the Stress Score is between 0 and 21.<sup>14</sup> The Stress Score counts categories rather than diagnoses, as it is in particular the heterogeneity of functional disorders (i.e. different categories) that highly correlates with stress (Eliassen et al., 2018). It is important to keep in mind that the Stress Score does not capture the full extent of stress as some individuals might not visit a physician in case of stress.

In our main analysis sample, the average value of the Disease based Stress Score is 1.21, the standard deviation is 1.36. Figure 2 provides an overview of the distribution of the Stress Score for our analysis sample. A unit increase in the Stress Scale is associated with a higher utilization of ambulatory health services by about one quarter among

---

<sup>12</sup>These subscales include cardiac arrhythmia; minor hypertension, cardiovascular problems; sleep disorder; sexual dysfunction; gastrointestinal disorders; eating disorders; head ache; back pain; minor depressive episode; minor anxiety disorders; adjustment disorders; mild cognitive disorder and concentration disorder; tinnitus, idiopathic hearing/voice loss, vertigo; subjective visual disturbances; immune dysfunction and allergy; diseases of the skin; alcohol abuse; somatoform disorders; problematic social situation; behavioral and emotional disorders; disorders of psychological development, learning disorders. Each subscale includes only diagnoses of low severity and does not consider serious diseases.

<sup>13</sup>Due to data privacy, we do not have access to the individual ICD-10 based diagnoses.

<sup>14</sup>An updated version of the Stress Score, called the Affected Mental and Physical Systems Index (AMPSI), was developed by Amélie Rouche at Charité - University Hospital Berlin. It excludes the subscale hypertension – a diagnosis that only 0.5 percent of children in our sample received - and, hence, ranges only from 0 to 20. Excluding hypertension does not change our results.

children (Rouche et al., 2017a). We do not only take the overall Stress Score as outcome, but we also consider binary variables indicating whether the Stress Score is larger than 0, 1, 2, and 3.

### 3.3 Empirical Strategy

Our empirical strategy exploits variation in time and across states in a difference-in-differences framework to identify the effect of the G8 reform on stress-related diagnoses. We have two sources of variation in time in our data: variation across years of observation (as used in previous studies on this reform) as well as variation across assigned grades (within year of observations). The latter also represents variation over time as, in a given data year, the grade depends on a child’s year and month of birth. In our main specification, we pool these two sorts of variation across time, while we show in the robustness section that our conclusions are robust to using only one source of variation across time. In our main specification, we estimate the following equation by OLS:

$$y_{igt} = \beta G8_{gt} + \lambda_t + \mu_g + \kappa_d + \varepsilon_{igt}, \quad (1)$$

where  $y_{igt}$  denotes the value of the Disease based Stress Score for individual  $i$  in grade  $g$  in county  $d$  and in year  $t$ ,  $G8_{gt}$  indicates whether grade  $g$  in year  $t$  and county  $d$  was subject to the G8 reform or not and  $\beta$  is the coefficient of interest, providing the effect of the G8 reform on the Stress Score.  $\kappa_d$  is a set of county fixed effects that takes general differences between counties into account, and, hence, also between states as each county belongs to a single state. These county fixed effects also take into account time-invariant regional differences in the coding behavior of physicians.  $\lambda_t$  denotes year fixed effects and  $\mu_g$  grade fixed effects. The latter two sets of fixed effects control for general differences across years and grades, respectively.  $\varepsilon_{igt}$  is the error term. In our main analysis, we allow for correlation of error terms within cohort-state cells, where a cohort comprises all students assigned to the same school entry year. Cohort is a linear combination of grade and observation year in our data. The cohort-state level is the level at which the treatment indicator is determined, which is a common level for clustering. In the robustness section, we show that our results are robust to alternative ways of clustering.

## 4 Results

### 4.1 Main results

Table 2 presents our estimates for the effect of the G8 reform on stress-related diagnoses. Controlling for year, county, and grade fixed effects the estimates in Panel A suggest that the Stress Score increased by 0.024 units following the introduction of the G8 reform. This statistically significant effect amounts to about 1.7 % of a standard deviation of the Stress Score and is similar to the difference between the Stress Score of counties at the 50th and the 56th percentile in our sample. This suggests that the effect is rather small. In column (2), we add county-specific linear trends to further strengthen the plausibility of the identification assumption. Next, we control for several time-varying regional variables in column (3). These variables include (i) the numbers of psychologists, pediatricians, and general practitioners per 100,000 individuals at the county level, taking into account changes in the supply of doctors that might affect our outcomes; (ii) the unemployment rate at the county-level, proxying for possible changes in the demand for mental health care; and (iii) annual expenses per student, taking into account other changes in the education system. Controlling for all these variables does not change our conclusions.<sup>15</sup> The other panels show that stress-related diagnoses increase at both the extensive and the intensive margin. The probability to have at least one stress-related diagnosis increased by 0.6 percentage points due to the G8 reform (Panel B). Compared to the sample share of 62.3 %, this amounts to an increase of about 1 %. Panels C, D, and E show that the G8 reform also significantly increased the probability that the Stress Score is larger than 1, 2, and 3, respectively.<sup>16</sup>

When further disentangling the results and taking the 21 subscales of the Stress Score as outcome variables, we see that no specific subscale drives the results.<sup>17</sup> Instead, the G8 reform increases several subscales. In absolute terms, the strongest effects are found for eating disorders, headaches, and gastrointestinal disorders. However, several other subscales also exhibit significant increases. Eating disorders are also the subscale for which we obtain the largest relative increase: The reform-induced an increase in eating disorders of 0.7 percentage points, which amounts to an increase of about 11 %, relative to the sample share of 6.4 % with eating disorders.

---

<sup>15</sup>A falsification test in Table A.2 also shows that the G8 reform is unrelated to the supply of physicians. The regressions in this table are based on equation (1) but use different outcome variables. None of the G8 coefficients are statistically significant.

<sup>16</sup>When looking at the probability that the Stress Score is larger than 4 (not shown), we find a small but borderline significant effect. There is no evidence that the G8 reform also increased the probability that the Stress Score exceeds 5 (not shown). However, it is important to note that only few individuals in our sample have a Stress Score value larger than 5 (1.2 %).

<sup>17</sup>Results available upon request from the authors.

Looking at the effects separately by grades, Table 3 shows that all the estimated coefficients are positive suggesting that the G8 reform increased stress-related diagnoses across grades. However, the effect on the Stress Score is only statistically significant in grades 9 and 10, the grades that received the strongest increases in instruction time. While the slightly smaller coefficients in grade 10 might suggest that students learn to cope with stress, the difference in the coefficient estimates between grade 9 and 10 is not statistically significant.<sup>18</sup>

## 4.2 Robustness

In this section, we provide additional results for the robustness of our findings. Our data set does not allow for distinguishing between individuals in different school tracks. As the G8 reform only affected the academic track, our estimates can be seen as lower bounds of the true reform effect. For the estimations in column (2) of Table 4, we scale the treatment effect by the share of treated students in a state- year-grade cell. For this purpose, we aggregate the data at the state- year-grade level and interact the dummy G8-treatment indicator in equation (1) with the share of academic-track students in a state- year-grade cell.<sup>19</sup> Unsurprisingly, this exercise suggests that the true effects are much larger than the results presented above because we cannot restrict our analysis to academic track students, instead estimating intent-to-treat effects.<sup>20</sup> However, even these scaled-up estimates remain quantitatively rather small: For instance, the increase in the Stress Score of 0.066 units still amounts to less than 5 % of a standard deviation.

The second robustness check (column 3) presents the results for a placebo policy change that took place one year before the actual G8 reform.<sup>21</sup> The effect of this placebo reform is small and statistically insignificant, suggesting that treatment and control

<sup>18</sup>Such a coping mechanism is well established in the context of *resilience* in the psychological literature, in particular in the *challenge model* of resilience (Zolkoski and Bullock, 2012; Masten, 2016).

<sup>19</sup>We also control for the main effect of the share of academic-track students. Aggregation for the scaling estimation is performed at the state level, which does not alter the point estimates, as we do not include any individual level information in the regressions and weight the state-level observations by the cell size see (see Angrist and Pischke, 2008).

<sup>20</sup>In these computations we basically assume a treatment effect of exactly zero for students in all other tracks. With an average rate of 36.6 % of academic track students, a first attempt of constructing an upper bound for the average treatment effect would be  $\hat{\beta}^{ITT} \times (1/0.366) = \hat{\beta}^{ITT} \times 2.73$ , which is close to the ratio of the coefficients from the main specification and the scaled coefficients. However, the scaling factor for the upper bound might be larger since not all students followed the school entry age cut-offs and there is grade repetition and grade skipping. If we reproduce our grade-assignment mechanism with the KiGGS survey data, we correctly assign the grade for 78 % of academic track students (see Appendix Table A.1). If we assume that all the wrong assignments affected G8 status, we obtain a larger upper bound for the average treatment effect:  $\hat{\beta}^{ITT} \times (1/(0.366 \times 0.78)) = \hat{\beta}^{ITT} \times 3.5$ .

<sup>21</sup>Column 3 is based on equation (1) but additionally includes an indicator variable for the placebo policy reform that took place one year before. The displayed coefficient refers to the placebo reform indicator.

group did not exhibit differential trends before the introduction of the G8 reform. This supports the identification assumption of common trends.

The next set of robustness checks deals with alternative sample restrictions. First, we restrict the sample to only West German states, as in our observation period it is mainly West German states that switched from G9 to G8. Second, we extend the sample to include grades 5-12 (i.e., all the grades at academic track schools under the G8 regime). Third, we exclude the first G8 cohort in order to investigate whether the effects are mainly driven by transition effects. Fourth, we exclude students from the last G9 cohort. An erroneous assignment of the grade due to late school enrollment or grade retention for individuals in this cohort, would result in the wrong assignment of the treatment indicator and, hence, a bias towards zero of our point estimates. Fifth, while our health claims data of the German Social Health Insurance only include individuals with at least one doctor visit in a given year, we merge information on the size of the respective age group in a given county. We then extend our sample by the “missing observations” and assume that their value of the Stress Score is equal to 0.<sup>22</sup> Sixth, we include also the state of Hesse, which implemented the reform across three cohorts.<sup>23</sup> The results in columns (4) to (9) in Table 4 show that our findings are robust to these alternative sample definitions.

In column (10), we replace the binary reform indicator with the number of weekly instruction hours. As this measure also considers changes in instruction time that were not induced by the G8 reform (and is, hence, more noisy), column (11) instead only studies the increase in instruction hours induced by the G8 reform.<sup>24</sup> Both specifications confirm our conclusions. The last set of robustness checks address various estimation issues. As our main outcome variable, the Stress Score, is a count variable, we apply specific count data models: Poisson regression (column 12)) and Negative Binomial regression (column 13).<sup>25</sup> In column (14), we cluster the standard errors on the state-grade level. Again, the results are robust to these alternative specification.

The next set of robustness checks deals with the different sources of variation in our data. As our data contains variation across two time dimensions (observation years

---

<sup>22</sup>Note that this is not our preferred specification. Our sample covers on average more than 85 % of a county’s population and the remaining 15 % are not only individuals from the SHI system without doctor visit in a given year but also individuals with private health insurance who are not part of the SHI system (and who might have stress-related diagnoses as well). In our observation period, about 70 million individuals were insured in the German SHI system (German Federal Ministry of Health, 2018), meaning that across all age groups the SHI system covers about 87 % of the population.

<sup>23</sup>In Hesse, 10 % of schools changed to the G8 system for the first of these cohorts, 60 % for the second cohort, and 30 % for the third cohort. We assign the treatment indicator based on the middle cohort.

<sup>24</sup>Technically, we replace the binary treatment indicator by the difference in weekly instruction hours between the first G8 cohort and the last G9 cohort.

<sup>25</sup>The columns for the Poisson and Negative Binomial model display average marginal effects.

and grades) as well as variation across states, the data allow three ways to set up a DiD design. First, exploiting only variation across states and years (holding grade constant). This means, we compare outcomes across states and years with and without G8 - within a grade. Second, exploiting only variation across states and grades (holding observation year constant). Third, exploiting only variation across grades and observation years (holding state constant). While we pool these sources of variation in our main specification, the last set of robustness checks shows that our conclusions do not change when we exploit only one source of variation. More specifically, in column (2) of Table 5, we exploit only the first type of variation, i.e. variation across states and years. For this specification, we estimate a version of equation (1) that includes fixed effects for each county-grade combination as well as for each year-grade combination. Similarly, in column (3), we exploit only the second type of variation by including fixed effects for each year-grade and county-year combination, thus holding observation year constant. Finally, column (4) includes fixed effects for each county-grade and county-year combination, thereby holding constant state/county and exploiting only the third type of variation. It is reassuring to see that the conclusions do not change when we consider only one source of variation.

As a last robustness exercise, we follow Anderson and Meyer (2000) and Conley and Taber (2011) and provide additional randomization inference. We randomly assign placebo G8 status to the different federal states such that the pattern of placebo G8 introductions follows the actual reform. For all five dependent variables of our main specification (Table 2, Column 1), we estimate the distribution of placebo G8 coefficients from 5000 random reform assignments. Results from this randomization inference exercise presented in Figure A.1 show that the actual G8 coefficients are in the 91st percentile of placebo coefficients for the stress scale and around the 90th percentile for the linear probability models for stress scales larger than 0, 1, 2, and 3. The results from this simulation suggest that the actual coefficients are unlikely to be observed if there was no G8 effect.<sup>26</sup>

## 5 Conclusion

This study contributes to our understanding of the interrelation between schooling and stress-related mental health problems among school children. More specifically, this study investigates the mental health consequences of a German schooling reform that substantially increased instruction time. Based on a difference-in-differences frame-

---

<sup>26</sup>Similar applications of randomization inference on the size of the coefficient in similar settings are also conducted by, among others, Gruber and Hungerman (2008) and Bullinger (2019).

work and stress-related outpatient diagnoses from the universe of health claims data of the German Social Health Insurance system, we find evidence that more instruction time slightly increases stress-related health problems of school children. The effects are robust to alternative model specifications.

This study contributes to the existing literature (i) by relying on a measure of stress-related mental health problems that is reported by physicians and not by the affected individuals themselves; (ii) by focusing on the identification of causal effects; and (iii) by looking at a specific element of the schooling system. When interpreting the results of our study, several limitations have to be kept in mind. First, while an advantage of our outcome measure is that it is not reported by the affected individuals themselves, it might not be able to capture the full extent of additional stress induced by more instruction time, as some individuals might not visit a physician in case of stress. However, our outcome measure is clearly less extreme than (attempted) suicide, a measure used in previous research on schooling and health. Second, the data from the German Social Health Insurance system does not include information about the school track and the school grade, so that there is some measurement error in the assignment of the treatment indicator and, hence, our estimates can be seen as lower bounds.

An important question that we cannot answer with the data at hand is whether the observed increases in stress due to increases of instructional time have long-term consequences on health or other later-life outcomes. Extensive literature indicates that experiencing extreme events during childhood, such as destruction or hunger during World War II has long-term health, educational and labor market consequences for the affected individuals (e.g. Akbulut-Yuksel, 2014; Kesternich et al., 2014). Furthermore, the association of adverse childhood experiences and later-life health is extensively documented in the epidemiological and public health literature, although the documented effects may not be causal (see Hughes et al. (2017) for an overview and discussion). Thus, we view the investigation of longer-term causal effects of less extreme stressors during childhood, such as increases in instructional time, as a fruitful area for future research.

Our findings also do not mean that the G8 reform and increasing weekly instruction time are bad education policies. Previous studies find that more instruction time improves student performance (e.g. Marcotte, 2007; Sims, 2008; Herrmann and Rockoff, 2012; Carlsson et al., 2015) and there is also evidence that the G8 reform increased academic performance in grade 9 due to the additional weekly instruction time (Huebener et al., 2017). Moreover, G8 students are younger when they graduate from high school (Huebener and Marcus, 2017) and when they enroll in university (Marcus and Zambre, 2019), thereby allowing for an earlier labor market entry. These beneficial effects of the



G8 reform have to be weighed against its adverse effects. In particular, previous studies find increased grade repetition rates (Huebener and Marcus, 2017), slightly lower performance at the end of schooling (Büttner and Thomsen, 2015; Huebener and Marcus, 2017; Huebener et al., 2017), and lower enrollment rates at university (Marcus and Zambre, 2019) as consequences of the G8 reform.

Taken together, our results carry important insights both from public health and education perspectives. From a public health perspective, our results support the notion that school can cause stress in children. In general, we know very little whether and how schooling affects the mental health of children. More research is needed on which specific elements of the schooling system particularly affect children's mental health.

From an education perspective our results therefore caution against pushing too hard to increase instruction time. Although – as discussed above - earlier literature shows that more instruction time can positively affect student performance, the improved student performance must be weighed against the additional stress students experience.

## References

- Akbulut-Yuksel, M. (2014). Children of war the long-run effects of large-scale physical destruction and warfare on children. *Journal of Human Resources*, 49(3):634–662.
- Anderson, P. M. and Meyer, B. D. (2000). The effects of the unemployment insurance payroll tax on wages, employment, claims and denials. *Journal of Public Economics*, 78(1-2):81–106.
- Ang, R. P. and Huan, V. S. (2006). Relationship between academic stress and suicidal ideation: Testing for depression as a mediator using multiple regression. *Child Psychiatry and Human Development*, 37(2):133.
- Angrist, J. D. and Pischke, J.-S. (2008). *Mostly harmless econometrics: An empiricist's companion*. Princeton University Press.
- Baltimore Center for Adolescent Health (2006). Confronting teen stress: Meeting the challenge in Baltimore City. Technical report, Baltimore Center for Adolescent Health.
- Banks, J. and Smyth, E. (2015). ‘Your whole life depends on it’: Academic stress and high-stakes testing in Ireland. *Journal of Youth Studies*, 18(5):598–616.
- Bokhari, F. and Schneider, H. (2011). School accountability laws and the consumption of psychostimulants. *Journal of Health Economics*, 30(2):355–372.
- Bor, W., Dean, A. J., Najman, J., and Hayatbakhsh, R. (2014). Are child and adolescent mental health problems increasing in the 21st century? A systematic review. *Australian & New Zealand Journal of Psychiatry*, 48(7):606–616.
- Bullinger, L. R. (2019). The effect of paid family leave on infant and parental health in the United States. *Journal of Health Economics*, 66:101–116.
- Büttner, B. and Thomsen, S. L. (2015). 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.
- Carlsson, M., Dahl, G. B., Öckert, B., and Rooth, D.-O. (2015). The Effect of Schooling on cognitive Skills. *The Review of Economics and Statistics*, 97(3):533–547.
- Collishaw, S. (2015). Annual Research Review: Secular Trends in Child and adolescent mental Health. *Journal of Child Psychology and Psychiatry*, 56(3):370–393.
- Collishaw, S., Maughan, B., Goodman, R., and Pickles, A. (2004). Time Trends in adolescent mental Health. *Journal of Child Psychology and Psychiatry*, 45(8):1350–1362.
- Conley, T. G. and Taber, C. R. (2011). Inference with “difference in differences” with a small number of policy changes. *The Review of Economics and Statistics*, 93(1):113–125.

- Conner, J. O., Miles, S. B., and Pope, D. C. (2014). How many Teachers does it take to support a Student?: Examining the Relationship between Teacher Support and adverse Health Outcomes in high-performing, pressure-cooker High Schools. *The High School Journal*, 98(1):22–42.
- Dahmann, S. and Anger, S. (2014). The Impact of Education on Personality: Evidence from a German High School Reform. *IZA Discussion Paper No. 8139*.
- Dahmann, S. C. (2017). How does Education improve cognitive Skills? Instructional Time versus Timing of Instruction. *Labour Economics*, 47:35–47.
- de Anda, D., Baroni, S., Boskin, L., Buchwald, L., Morgan, J., Ow, J., Gold, J. S., and Weiss, R. (2000). Stress, Stressors and Coping among High School Students. *Children and Youth Services Review*, 22(6):441–463.
- Die Welt (2016). Kurze Schulzeit, mehr Stunden – der G8-Stress. <http://www.welt.de/regionales/hamburg/article108288244/Kurze-Schulzeit-mehr-Stunden-der-G8-Stress.html>, (accessed on September 17th, 2018).
- Elias, M. J. (1989). Schools as a Source of Stress to Children: An Analysis of causal and ameliorative Influences. *Journal of School Psychology*, 27(4):393–407.
- Eliassen, M., Schröder, A., Fink, P., Kreiner, S., Dantoft, T. M., Poulsen, C. H., Petersen, M. W., Eplov, L. F., Skovbjerg, S., and Jørgensen, T. (2018). A step towards a new delimitation of functional somatic syndromes: A latent class analysis of symptoms in a population-based cohort study. *Journal of Psychosomatic Research*, 108:102–117.
- Eriksen, T. L. M., Nielsen, H. S., and Simonsen, M. (2014). Bullying in elementary school. *Journal of Human Resources*, 49(4):839–871.
- F.A.Z. (2014). Warum ist G8 gescheitert? 23.02.2014, page 61.
- Felitti, V. J., Anda, R. F., Nordenberg, D., Williamson, D. F., Spitz, A. M., Edwards, V., Koss, M. P., and Marks, J. S. (2019). Relationship of childhood abuse and household dysfunction to many of the leading causes of death in adults: The Adverse Childhood Experiences (ACE) Study. *American Journal of Preventive Medicine*, 56(6):774–786.
- Freeman, J. G., Samdal, O., Băban, A., and Bancila, D. (2012). The Relationship between School Perceptions and psychosomatic Complaints: Cross-country Differences across Canada, Norway, and Romania. *School Mental Health*, 4(2):95–104.
- Galloway, M., Conner, J., and Pope, D. (2013). Nonacademic Effects of Homework in privileged, high-performing High Schools. *The Journal of Experimental Education*, 81(4):490–510.

- German Federal Ministry of Health (2018). Gesetzliche Krankenversicherung – Kennzahlen und Faustformeln. Accessed: Januar 2018.
- German Federal Statistical Office (2012). Allgemeinbildende Schulen: Fachserie 11, Reihe 1. <https://www.destatis.de/DE/Publikationen/Thematisch/BildungForschungKultur/Schulen/BroschuereSchulenBlick.html>.
- German Federal Statistical Office (2014). *Bildungsausgaben - Ausgaben je Schülerin und Schüler 2011*. German Federal Statistical Office, Wiesbaden.
- Gruber, J. and Hungerman, D. M. (2008). The church versus the mall: What happens when religion faces increased secular competition? *The Quarterly Journal of Economics*, 123(2):831–862.
- Hansen, B. and Lang, M. (2011). Back to School Blues: Seasonality of Youth Suicide and the academic Calendar. *Economics of Education Review*, 30(5):850–861.
- Heissel, J. A., Adam, E. K., Doleac, J. L., and Figlio, D. N. (2018). Testing, stress, and performance: How students respond physiologically to high-stakes testing. *NBER Working Paper*, 25305.
- Herrmann, M. A. and Rockoff, J. E. (2012). Worker Absence and Productivity: Evidence from Teaching. *Journal of Labor Economics*, 30(4):749–782.
- Hofmann, S. and Mühlenweg, A. (2018). Learning Intensity Effects in Students’ mental and physical Health – Evidence from a large Scale natural Experiment in Germany. *Economics of Education Review*, 67:216–234.
- Hübner, N., Wagner, W., Kramer, J., Nagengast, B., and Trautwein, U. (2017). Die G8-Reform in Baden-Württemberg: Kompetenzen, Wohlbefinden und Freizeitverhalten vor und nach der Reform. *Zeitschrift für Erziehungswissenschaft*, 20(4):748–771.
- 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. (2015). Empirische Befunde zu Auswirkungen der G8-Schulzeitverkürzung. *DIW Roundup*, 57.
- Huebener, M. and Marcus, J. (2017). Compressing Instruction Time into fewer Years of Schooling and the Impact on Student Performance. *Economics of Education Review*, 58:1–18.
- Hughes, K., Bellis, M. A., Hardcastle, K. A., Sethi, D., Butchart, A., Mikton, C., Jones, L., and Dunne, M. P. (2017). The effect of multiple adverse childhood experiences on health: a systematic review and meta-analysis. *The Lancet Public Health*, 2(8):e356–e366.

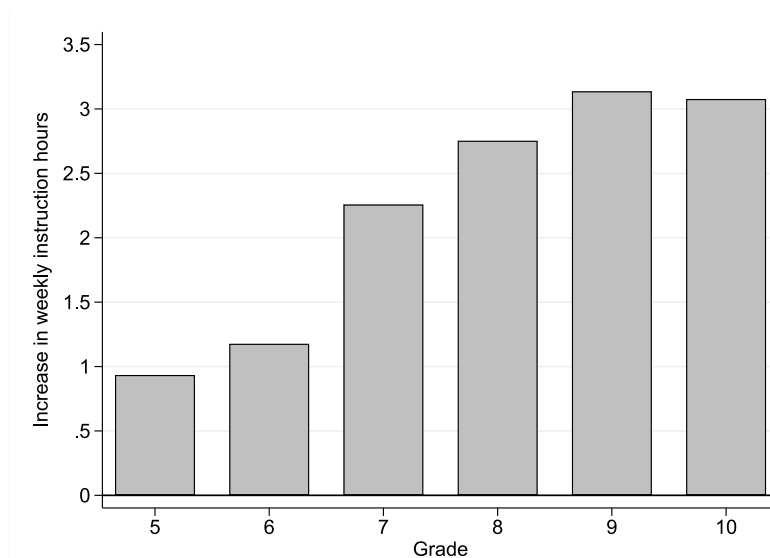
- Jewett, R., Sabiston, C. M., Brunet, J., O'Loughlin, E. K., Scarapicchia, T., and O'Loughlin, J. (2014). School Sport Participation during Adolescence and mental Health in early Adulthood. *Journal of Adolescent Health*, 55(5):640–644.
- Kesternich, I., Siflinger, B., Smith, J. P., and Winter, J. K. (2014). The Effects of World War II on Economic and Health Outcomes across Europe. *The Review of Economics and Statistics*, 96(1):103–118.
- King, M. D., Jennings, J., and Fletcher, J. M. (2014). Medical Adaptation to academic Pressure: Schooling, Stimulant use, and socioeconomic Status. *American Sociological Review*, 79(6):1039–1066.
- Lahti, A., Räsänen, P., Karvonen, K., Särkioja, T., Meyer-Rochow, V. B., and Hakko, H. (2005). Autumn peak in Shooting Suicides of Children and Adolescents from Northern Finland. *Neuropsychobiology*, 54(2):140–146.
- Lechner, M., Rodriguez-Planas, N., and Fernández Kranz, D. (2016). Difference-in-difference estimation by FE and OLS when there is panel non-response. *Journal of Applied Statistics*, 43(11):2044–2052.
- Leung, G. S., Yeung, K. C., and Wong, D. F. (2010). Academic Stressors and Anxiety in Children: The Role of paternal Support. *Journal of Child and Family studies*, 19(1):90–100.
- Löfstedt, P. (2017). Why are psychosomatic Symptoms in young People increasing in Sweden? *European Journal of Public Health*, 27(Supplement 3):388.
- Marcotte, D. E. (2007). Schooling and Test Scores: A mother-natural Experiment. *Economics of Education Review*, 26(5):629–640.
- Marcus, J. and Zambre, V. (2019). The Effect of Increasing Education Efficiency on University Enrollment: Evidence from administrative Data and an unusual Schooling Reform in Germany. *Journal of Human Resources*, 54(2):468–502.
- Masten, A. S. (2016). Resilience in developing systems: The promise of integrated approaches. *European Journal of Developmental Psychology*, 13(3):297–312.
- Matsubayashi, T., Ueda, M., and Yoshikawa, K. (2016). School and Seasonality in Youth Suicide: Evidence from Japan. *Journal of Epidemiology and Community Health*, 70(11):1122–1127.
- Meyer, T. and Thomsen, S. L. (2015). Schneller fertig, aber weniger Freizeit? – Eine Evaluation der Wirkungen der verkürzten Gymnasialschulzeit auf die außerschulischen Aktivitäten der Schülerinnen und Schüler. *Schmollers Jahrbuch*, 135(3):249–277.
- Meyer, T. and Thomsen, S. L. (2016). How important is Secondary School Duration for post-school Education Decisions? Evidence from a natural Experiment. *Journal of Human Capital*, 10(1):249–278.

- 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(04):255–260.
- Quis, J. S. (2018). Does compressing High School Duration affect Students’ Stress and mental Health? Evidence from the National Educational Panel Study. *Journal of Economics and Statistics*, 235(5).
- Quis, J. S. and Reif, S. (2017). Health Effects of Instruction Intensity – Evidence from a natural Experiment in German High-Schools. *FAU Discussion Papers in Economics*, 12-2017.
- Ravens-Sieberger, U., Freeman, J., Kokonyei, G., Thomas, C. A., and Erhart, M. (2009). School as a Determinant for Health Outcomes – A structural Equation Model Analysis. *Health Education*, 109(4):342–356.
- RKI (2015). The german health survey for children and adolescents (kiggs wave 1). public use file. *Robert Koch Institute, Department of Epidemiology and Health Monitoring*.
- Rouche, A., Erhart, M., and Graf von Stillfried, D. (2017a). Stress-Index anhand ambulanter Routine-Versorgungsdaten – ein Werkzeug zur individuellen Schätzung anhaltender psychischer Belastung. <http://doi.org/10.1055/s-0037-1606018>. *Das Gesundheitswesen*, 79:656 – 804.
- Rouche, A., Erhart, M., and von Stillfried, D. (2017b). A Stress Index based on ambulatory Claims Data for the Detection of mental Causes of Health Complaints. In *WPA XVII World Congress of Psychiatry (2017 Oct 8-12; Berlin, Germany)*.
- Schwandt, H. and Wuppermann, A. (2016). The Youngest get the Pill: ADHD Misdiagnosis in Germany, its regional Correlates and international Comparison. *Labour Economics*, 43:72–86.
- Sims, D. P. (2008). Strategic Responses to School Accountability Measures: It’s all in the Timing. *Economics of Education Review*, 27(1):58–68.
- Song, H. (2017). The Cost of attending an Elite College. *Economics Letters*, 159:173–176.
- Spiegel Online (2016). Streit über G8: Bayerns Schulminister erklärt Turbo-Abitur für „überholt“. <http://www.spiegel.de/schulspiegel/wissen/abitur-g8-spaenle-nennt-turbo-gymnasium-ueberholt-a-959660.html>, (accessed on September 17th, 2018).
- 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. and Anger, S. (2018). Die Notwendigkeit ökonomischer Politikberatung für eine evidenzbasierte Bildungspolitik: Verkürzung und Verlängerung der Schulzeit am Gymnasium. *Perspektiven der Wirtschaftspolitik*, 19(3):167–184.

- Torsheim, T. and Wold, B. (2001). School-related Stress, School Support, and somatic Complaints: A general Population Study. *Journal of Adolescent Research*, 16(3):292–303.
- Vossenkuhl, A. (2010). (Berufs-)Schulpflicht in Deutschland. *Berufsbildung in Wissenschaft und Praxis*, 39(6):53–54.
- WHO (2016). *Growing up unequal: gender and socioeconomic differences in young people's health and well-being. Health behaviour in school-aged children (HBSC) study: international report from the 2013/2014 survey*. World Health Organization, Copenhagen.
- Zolkoski, S. M. and Bullock, L. M. (2012). Resilience in children and youth: A review. *Children and Youth Services Review*, 34(12):2295–2303.

## Figures and Tables

Figure 1: Increase in instruction hours due to reform



*Notes:* Grade specific changes in weekly instruction hours due to the G8 reform. Adapted from Huebener et al. (2017). The average increases are estimated from official timetable regulations for our estimation sample.

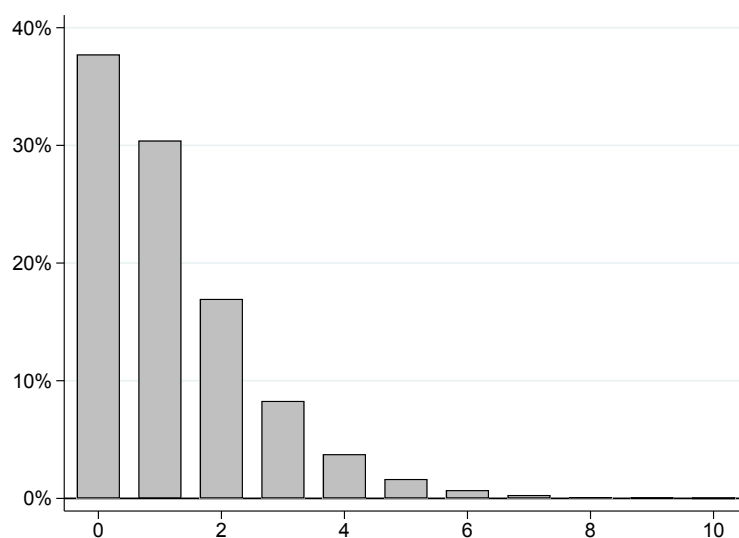


Table 1: Overview over the implementation of the G8 reform across German states

	First affected cohort in grade 7	First affected birth cohort	First G8 graduates
<b>Change from G9 to G8</b>			
Saxony-Anhalt (ST)	2001	06/1988	2007
Mecklenburg-Vorpommern (MV)	2002	06/1989	2008
Saarland (SL)	2003	07/1990	2009
Hamburg (HH)	2004	07/1991	2010
Bavaria (BY)	2005	07/1992	2011
Lower-Saxony (NI)	2005	07/1992	2011
Baden-Württemberg (BW)	2006	07/1993	2012
Bremen (HB)	2006	07/1993	2012
Berlin (BE)	2006	07/1993	2012
Brandenburg (BB)	2006	06/1993	2012
Hesse (HE)*	2006	07/1993	2012
North Rhine-Westphalia (NW)	2007	07/1994	2013
Schleswig-Holstein (SH)	2010	07/1997	2016
<b>Always G8</b>			
Saxony (SN)	all	all	all
Thuringia (TH)	all	all	all
<b>Always G9</b>			
Rhineland-Palatinate (RP)	none	none	none

*Notes:* Overview over the introduction of the G8 reform across German states. Several states have announced that they will abolish the G8 reform and switch back to the old system with 13 years of schooling, including Lower Saxony (last G8 graduates in 2019), Bavaria (last G8 graduates in 2024), North Rhine-Westphalia (last G8 graduates in 2024), and Schleswig-Holstein (last G8 graduates in 2025). In Hesse, the majority of academic track schools abolished the G8 reform at different points in time. All switch-backs take place after our observation period. Hesse implemented the G8 reform across the three school entry cohorts from 07/1993-07/1995: 10 % of schools changed to the G8 system for the first of these cohorts, 60 % for the second cohort, and 30 % for the third cohort.

Figure 2: Distribution of the Disease based Stress Score (DbSS)



*Notes:* The distribution of the Stress Score in our sample. Source: Own calculation based on German health insurance claims data.

Table 2: Effect of the G8 reform on the Stress Score

	Mean	(1)	(2)	(3)
<b>Panel A: Stress Scale</b>				
G8 reform	1.208	0.024** (0.009)	0.018*** (0.006)	0.025*** (0.009)
<b>Panel B: Pr(Stress Scale &gt; 0)</b>				
G8 reform	.623	0.006** (0.003)	0.003* (0.002)	0.006** (0.003)
<b>Panel C: Pr(Stress Scale &gt; 1)</b>				
G8 reform	.319	0.007** (0.003)	0.004** (0.002)	0.007** (0.003)
<b>Panel D: Pr(Stress Scale &gt; 2)</b>				
G8 reform	.15	0.005** (0.002)	0.004*** (0.001)	0.005** (0.002)
<b>Panel E: Pr(Stress Scale &gt; 3)</b>				
G8 reform	.067	0.003*** (0.001)	0.003*** (0.001)	0.003*** (0.001)
N		10,056,859	10,056,859	10,056,859
County FE:		yes	yes	yes
Grade FE:		yes	yes	yes
Year FE:		yes	yes	yes
County trends		-	yes	-
Regional controls		-	-	yes

*Notes:* The tables displays the effect of the G8 reform on stress-related outpatient diagnoses. Regional controls include the unemployment rate at the county-level, the states' annual expenses per student, as well as the numbers of psychologists, pediatricians, and general practitioners, each per 100,000 individuals at the county level. Standard errors are clustered at the cohort-state level. \*  $p < 0.1$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ .

Table 3: Effect in different grades

	Stress	Stress Score larger than			
	Score	0	1	2	3
Grade 7	0.022 (0.013)	0.006 (0.005)	0.006 (0.004)	0.005** (0.003)	0.003* (0.002)
Grade 8	0.011 (0.014)	0.001 (0.005)	0.004 (0.004)	0.003 (0.003)	0.002 (0.001)
Grade 9	0.032*** (0.008)	0.009*** (0.003)	0.009*** (0.003)	0.007*** (0.002)	0.004*** (0.001)
Grade 10	0.023** (0.011)	0.005* (0.003)	0.006* (0.003)	0.005* (0.002)	0.003** (0.001)
N	10,056,859	10,056,859	10,056,859	10,056,859	10,056,859

*Notes:* The table presents the effects of the G8 reform on stress-related outpatient diagnoses in different grades. Each column comes from a separate regression that is based on equation (1) and includes interactions between the G8 treatment indicator and each grade. All regressions include county, year, and grade fixed effects. Standard errors are clustered at the cohort-state level. \*  $p < 0.1$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ .

Table 4: Robustness tests

	Main (1)	Scale- up (2)	Plac. reform (3)	Alternative samples						Alternative treatment def.		Estimation issues		
				Only West (4)	Grades 5-12 (5)	Excl. first G8 (6)	Excl. last G9 (7)	Full popul. (8)	With Hesse (9)	Instr. hours (10)	Δ Instr. hours (11)	Poisson (12)	Neg. bin. (13)	State-grade cluster (14)
Panel A: Stress Scale														
G8 reform	0.024** (0.009)	0.066** (0.028)	0.005 (0.015)	0.022** (0.009)	0.023*** (0.008)	0.031*** (0.010)	0.029* (0.017)	0.021** (0.009)	0.020** (0.009)	0.006* (0.003)	0.007*** (0.002)	0.024** (0.010)	0.024** (0.010)	0.024** (0.010)
Panel B: Pr(Stress Scale > 0)														
G8 reform	0.006** (0.003)	0.016** (0.008)	0.000 (0.003)	0.006** (0.003)	0.003 (0.002)	0.008*** (0.003)	0.006 (0.004)	0.006* (0.003)	0.005** (0.003)	0.001* (0.001)	0.001 (0.001)			0.006** (0.003)
Panel C: Pr(Stress Scale > 1)														
G8 reform	0.007** (0.003)	0.018** (0.008)	0.001 (0.004)	0.006** (0.003)	0.005** (0.002)	0.008** (0.003)	0.008* (0.004)	0.006** (0.003)	0.005* (0.003)	0.001 (0.001)	0.002** (0.001)			0.007** (0.003)
Panel D: Pr(Stress Scale > 2)														
G8 reform	0.005** (0.002)	0.014** (0.006)	0.001 (0.003)	0.004** (0.002)	0.006*** (0.002)	0.007*** (0.002)	0.006* (0.004)	0.004** (0.002)	0.004** (0.002)	0.001* (0.001)	0.002*** (0.000)			0.005** (0.002)
Panel E: Pr(Stress Scale > 3)														
G8 reform	0.003*** (0.001)	0.009** (0.003)	0.001 (0.002)	0.003** (0.001)	0.004*** (0.001)	0.004*** (0.001)	0.004 (0.002)	0.002** (0.001)	0.003** (0.001)	0.001** (0.000)	0.001*** (0.000)			0.003*** (0.001)
N	10,056,859	240	10,056,859	8,526,879	20,740,983	9,046,365	9,493,420	11,716,151	10,886,906	10,056,859	10,056,859	10,056,859	10,056,859	10,056,859

Notes: All regressions control for county, grade, and year fixed effects. Standard errors are clustered at the cohort-state level. \*  $p < 0.1$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ .

Table 5: Exploiting different DiD dimensions

	(1)	(2)	(3)	(4)
<b>Panel A: Stress Scale</b>				
G8 reform	0.024** (0.009)	0.018** (0.008)	0.010* (0.006)	0.008*** (0.003)
<b>Panel B: Pr(Stress Scale &gt; 0)</b>				
G8 reform	0.006** (0.003)	0.006*** (0.002)	0.002 (0.002)	0.002* (0.001)
<b>Panel C: Pr(Stress Scale &gt; 1)</b>				
G8 reform	0.007** (0.003)	0.005** (0.002)	0.003* (0.002)	0.002 (0.001)
<b>Panel D: Pr(Stress Scale &gt; 2)</b>				
G8 reform	0.005** (0.002)	0.003* (0.002)	0.002* (0.001)	0.001 (0.001)
<b>Panel E: Pr(Stress Scale &gt; 3)</b>				
G8 reform	0.003*** (0.001)	0.002* (0.001)	0.002* (0.001)	0.001*** (0.000)
N	10,056,859	10,056,859	10,056,859	10,056,859
County FE:	yes	yes	yes	yes
Grade FE:	yes	yes	yes	yes
Year FE:	yes	yes	yes	yes
CountyXGrade FE:	-	yes	-	yes
YearXGrade FE:	-	yes	yes	-
CountyXYear FE:	-	-	yes	yes

*Notes:* The tables displays the effect of the G8 reform on stress-related outpatient diagnoses exploiting different sources of variation. Standard errors are clustered at the cohort-state level. \*  $p < 0.1$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ .

## Appendix

Table A.1: Population share in SHI

	Share SHI	Share High School	Correct Grade	N
All aged 10–17	87.52 %	33.23 %	71.43%	4,770
All aged 12–15	86.39 %	36.57 %	72.59 %	2,554
High school students aged 12–15	77.59 %	100 %	78.00 %	1,216

*Notes:* Based on the KiGGS survey (RKI, 2015). Shares of individuals in the social health insurance (SHI) or academic track schools (High School). Correct Grade indicates whether the imputed grade from school-entry cutoff rules correspondent to the reported grade. Shares obtained using cross sectional weights.

Table A.2: Falsification test: "Effect" of the G8 reform on the supply of physicians

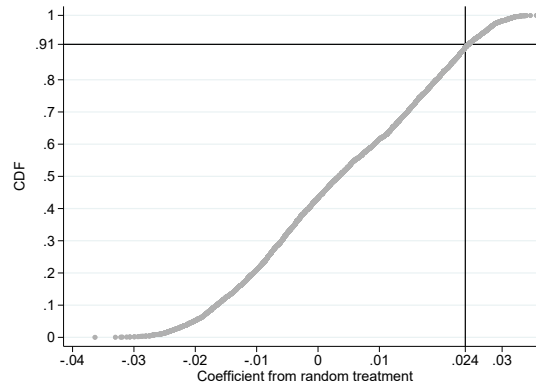
	Psychos (1)	Pedias (2)	GPs (3)
G8 reform	0.126 (0.099)	0.013 (0.014)	-0.010 (0.124)
N	10,056,859	10,056,859	10,056,859

*Notes:* The tables displays the effect of the G8 reform on the supply of physicians. All regressions are based on equation (1) but use different outcome variables. Outcome in is the number of psychologists per 100,000 inhabitants in (1), of pediatricians per 100,000 inhabitants in (2), and of GPs per 100,000 inhabitants in (3). Standard errors are clustered at the cohort-state level. \*  $p < 0.1$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ .

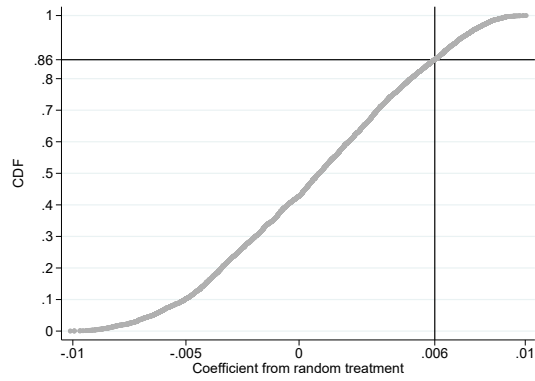


Figure A.1: Randomization Inference

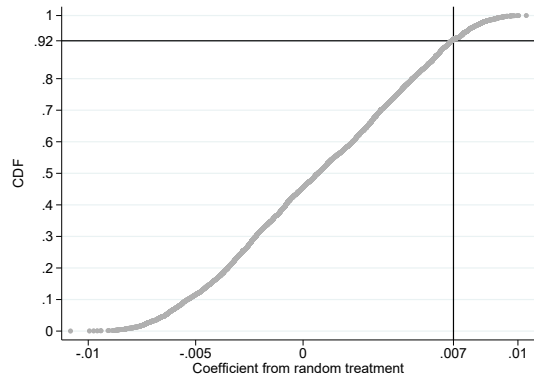
### Stress Scale



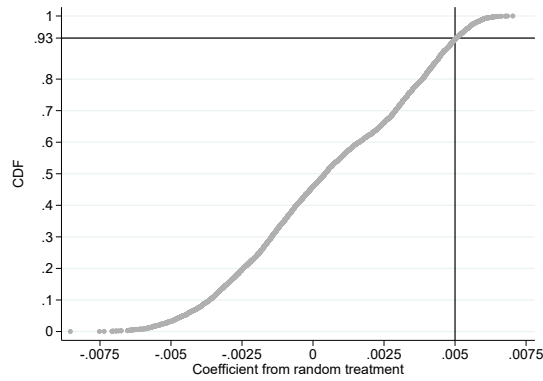
### Pr(Stress Scale > 0)



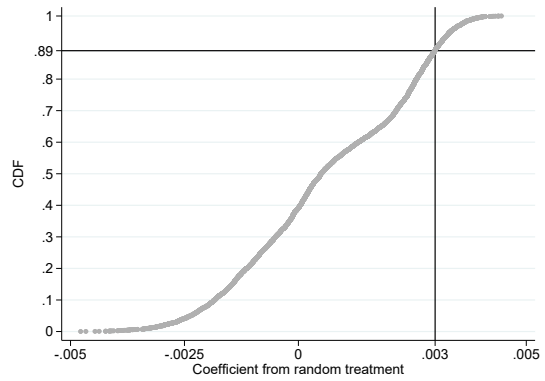
### Pr(Stress Scale > 1)



### Pr(Stress Scale > 2)



### Pr(Stress Scale > 3)



Notes: Cumulative distribution function (CDF) of coefficients from random assignment of treatment status for the main specification (Table 2, Column 1). Vertical lines indicate the real coefficient, horizontal lines indicate share of random coefficients smaller than the real one.