Schwerter, Jakob; Netz, Nicolai; Hübner, Nicolas

Does instructional time at school influence study time at university? Evidence from an instructional time reform

Economics of education review (2024) 100, 16 S.

Quellenangabe/ Reference:
Schwerter, Jakob; Netz, Nicolai; Hübner, Nicolas: Does instructional time at school influence study time at university? Evidence from an instructional time reform - In: Economics of education review (2024) 100, 16 S. - URN: urn:nbn:de:0111-pedocs-331352 - DOI: 10.25656/01:33135; 10.1016/j.econedurev.2024.102526

https://nbn-resolving.org/urn:nbn:de:0111-pedocs-331352
https://doi.org/10.25656/01:33135

Economics of Education Review 100 (2024) 102526

Contents lists available at ScienceDirect

Economics of Education Review

journal homepage: www.elsevier.com/locate/econedurev

Does instructional time at school influence study time at university? Evidence from an instructional time reform[☆]

Jakob Schwerter[a,*], Nicolai Netz[b], Nicolas Hübner[c]

[a] Center for Research on Education and School Development (IFS) & Faculty of Statistics, TU Dortmund University, Germany
[b] German Centre for Higher Education Research and Science Studies (DZHW), Germany
[c] Institute of Education, University of Tübingen, Germany

ARTICLE INFO

JEL classification:
I2
I21
I23

Keywords:
G8 high school reform
Higher education
Study time
Long-term effect
Staggered difference-in-differences

ABSTRACT

Early-life environments can have long-lasting effects on individuals' later life courses. Interestingly, research on the effects of school reforms has hardly adopted this perspective. Therefore, we investigate a staggered school reform that reduced the number of school years and increased weekly instructional time for secondary school students in most German federal states. We analyze this quasi-experiment in a difference-in-differences framework using nationally representative large-scale survey data on 69,513 students who attended university between 1998 and 2016. Using both TWFE and weighted-group ATT estimators, we find negative effects of reform exposure on hours spent attending classes and on self-study. Moreover, reform exposure increased the time gap between school completion and higher education entry. Our results support the view that research should examine unintended long-term effects of school reforms on individual life courses.

1. Introduction

Environments during childhood and adolescence influence a wide range of important outcomes in the later life course, such as skill development (Tymms et al., 2018), labor market performance (Gertler et al., 2014), and health (Taylor, 2010). Interestingly, very few studies have investigated whether policy reforms that change the learning environment have long-lasting effects on students' study time later in life. Such knowledge of potentially unintended long-term effects of policy interventions, however, is important for the planning, development, and implementation of future educational reforms.

To address this research gap, we exemplarily investigated whether a major school reform in Germany – known as the G8 reform – had long-term effects on students in higher education. The G8 reform reduced the overall duration of the academic-track lower secondary school (*Gymnasium*) from 9 to 8 years, while increasing weekly instructional time by about 3.68 h or 12.5 % (Hübner et al., 2022; Huebener et al., 2017), which led to reductions in students' leisure time (Hübner et al., 2017; Milde-Busch et al., 2010; Quis, 2018). Previous studies showed that the reform tended to lead to worse health outcomes (Dahmann & Anger, 2014; Hübner et al., 2017; Marcus et al., 2020; Quis, 2018) and mostly negative educational outcomes at secondary school (Büttner & Thomsen, 2015; Hübner et al., 2017; Huebener et al., 2017; Huebener & Marcus, 2017). First studies examining students in higher education also suggested that G8 students are somewhat more likely to delay higher education enrollment, drop out, and change their fields of study (Marcus & Zambre, 2019; Meyer et al., 2019).[1]

Building on these research findings, we tested whether the G8 reform also affected study time in higher education. Theoretically, students might have become accustomed to learning at a higher intensity in secondary school and maintained this intensity in higher education (habituation scenario). Conversely, they might also have reduced their

☆ We thank Adam Ayaita, Morgan Polikoff, Sebastian Lang, as well as two anonymous reviewers for their valuable feedback on prior versions of this manuscript.
* Corresponding author.
E-mail address: jakob.schwerter@tu-dortmund.de (J. Schwerter).

[1] Any educational reform is specific to the national context in which it takes place. However, reforms implemented in one country or state often serve as templates for reforms that are later implemented in other settings. An example is the District 2 reform in New York, which introduced balanced literacy and constructivist math programs and was later used as a blueprint for other city districts (Ravitch, 2011). In Germany, discussions about the optimal amount of instructional time are still ongoing: For instance, many German states (e.g., Bayern, Niedersachsen, and Schleswig-Holstein) have started to transition back to the G9 regime. Other states have recently allowed schools to choose between G8 and G9 regimes (Hessen and Nordrhein-Westfalen). These examples illustrate the ongoing relevance of examining educational effects of the G8 reform.

https://doi.org/10.1016/j.econedurev.2024.102526
Received 21 April 2023; Received in revised form 11 March 2024; Accepted 14 March 2024
Available online 9 May 2024

learning intensity in higher education to recover from their stressful school years or because they have acquired a negative attitude toward formal education (compensation scenario). Determining the effects of the G8 reform on students' study time is of high importance, as several studies have shown a causal link between time investment in one's studies and performance in higher education (Andrietti, 2014; Andrietti & Velasco, 2015; Arulampalam et al., 2012; Ersoy, 2021; Grave, 2011; Metcalfe et al., 2019; Schwerter et al., 2022a), which in turn improves labor market outcomes (Kittelsen Røberg & Helland, 2017). Therefore, our analysis of the long-term effects of secondary school reforms on study time at university closes an important research gap.

The different German federal states introduced the G8 reform on a staggered basis, leading to exogenous variation in instructional time across states. We exploited this quasi-experimental setting created by the staggered reform using a difference-in-differences (DiD) framework. Going beyond existing studies on the effects of the G8 reform, we estimated weighted group-time average treatment effects on the treated (ATT), as introduced by Callaway and Sant'Anna (2021b). This method overcomes several caveats of the two-way fixed effects (TWFE) method for examining time-staggered reforms. As a robustness check, however, we also present conventional TWFE estimates.

Using rich survey data on 69,513 university students in Germany, we found that G8 students spent less time attending classes and on self-study at university than G9 students. We did not find evidence that G8 students invested their freed-up time in additional work. Instead, G8 students tended to enroll in university after longer time gaps compared to G9 students. Thus, our study may explain why G8 students are more likely to drop out of higher education (Marcus & Zambre, 2019) despite their similar levels of intelligence compared to their peers (Dahmann, 2017): They devote less time to learning, both in classes and on their own. This finding is important because it strongly aligns with research indicating that students benefit from higher study time and effort (Metcalfe et al., 2019; Schwerter et al., 2022a, Schwerter et al., 2022b; Stinebrickner & Stinebrickner, 2008). Attending more classes (Grave, 2011) and, in particular, spending more time on self-study (Andrietti & Velasco, 2015; Arulampalam et al., 2012; Bonesrønning & Opstad, 2015; Bratti & Staffolani, 2013) improves exam performance. Notably, we also examined heterogeneity in the reform effects by gender and social background. Whereas the reform effects were very similar for first- and (at least) second-generation students (operationalization of social background), our results suggest that males were slightly more negatively affected by the reform than females. In summary, our results support the view that more intensive daily instruction can have adverse effects on children's development in the long run. On a broader note, they suggest that future research should more closely examine (un) intended long-term effects of school reforms on individual life courses.

This requires a deeper understanding of how early learning environments shape individuals' learning behavior in the long term. In this respect, different social science disciplines help us achieve this understanding: The sociological life course perspective highlights that experiences and decisions in the early life course create path dependencies that influence experiences and decisions in later life (Elder et al., 2003). As a result of their socialization throughout childhood and adolescence, individuals tend to accumulate resources depending on their parents' economic, social, and cultural capital (Bourdieu, 1973). The life course approach underscores that even slight differences in individuals' early-age resources and life chances may lead to severe social inequalities in later life (DiPrete & Eirich, 2006). Notably, such cumulative (dis)advantages may not only result from differences in individuals' early-age resource endowments; they are also shaped by the institutional settings that individuals pass through during the life course. Such institutional settings include the different stages of the education system, which may be abruptly altered by structural reforms (Mayer, 2004).

In line with these sociological propositions, psychological and school effectiveness research has shown how early learning environments shape student learning and motivation later in life. For instance, prior studies have found that the negative association between a reference group's achievement and individuals' perception of their own abilities (the big-fish-little-pond effect) has negative effects on achievement and aspirations that can still be detected five years after graduation from high school (Marsh & O'Mara, 2010). This line of inquiry also illustrated that being part of an effective learning environment at the beginning of school is associated with higher educational attainment more than ten years later (Tymms et al., 2018). Moreover, stressful childhood environments are closely tied to health risks in adulthood (Taylor, 2010).

Relatedly, research on the economics of education amply demonstrated how early changings to their school environments affect individuals' labor market outcomes later in life (Heckman et al., 2013). Abramitzky et al. (2021) showed that economic incentives changing children's schooling decisions in the early years of education increased later years of schooling and wages. Also, a free school choice program targeting disadvantaged students in Israel improved students' higher education enrollment in universities and teachers' colleges, as well as earnings later in life (Lavy, 2021). Moreover, changes in the curriculum, that is, intensified math instruction for low-ability students in lower secondary school, increased high school graduation rates and college enrolment (Cortes et al., 2015; Nomi & Raudenbush, 2016). Similarly, changing course selection opportunities in high school affected students' choices of study areas higher education (Biewen & Schwerter, 2022). Lastly, the literature highlights that more education (e.g., more years of compulsory schooling) lead to higher wages (Angrist & Krueger, 1991; Henderson et al., 2011; Oreopoulos, 2007), higher job prestige (Oreopoulos & Salvanes, 2011), and job satisfaction (Winkelmann & Winkelmann, 1998).

In summary, educational research from different disciplines illustrates that experiences and learning environments in the early life course can have long-lasting influences on later life. Hence, alterations in the early structure of learning environments may have a cascading impact on student learning behaviors in their future. Surprisingly, there is no established line of research testing this proposition. To narrow this research gap, we investigated the effects of an instructional time reform (the G8 reform) on students' study time in higher education.

1.1. Previous research on instructional time reforms

A long-standing debate addressed the importance of instructional time for student learning (Patall et al., 2010; Pischke, 2007). This debate produced rich, albeit mixed evidence – with some studies suggesting negative effects of more instructional time, some suggesting positive effects, and others suggesting that the effects vary depending on the examined student characteristics and outcomes (Allensworth et al., 2009; Andersen et al., 2016; Lavy, 2015; Meyer & Van Klaveren, 2013). Most of these studies investigated whether increased instructional time led to higher student achievement. Theoretically, these studies are in line with traditional school effectiveness models assuming that properties of the learning process (e.g., time on task) influence how specific learning inputs translate to specific learning outputs (Scheerens, 1990). They also align with theoretical frameworks such as the "Carroll Model of School Learning" (Carroll, 1988) or the concept of "Mastery Learning" (Bloom, 1968), which suggest that learning outcomes can be explained (and increased) by an optimal balance between students' aptitude and the amount of learning time.

Consistent with these models are the results of a reform in Ontario, Canada, which shortened college-preparatory high school by one year. Krashinsky (2014) and Morin (2013)) concluded that student achievement declined due to the reform, especially among lower-performing students. Thus, reducing instructional time alone led to a decline in achievement. This is in line with other findings that more years of schooling increase student achievement (Card, 1999; Lochner, 2011).

In recent decades, most German federal states have introduced a reform compressing school time (Hübner et al., 2022). Unlike the reform in Canada, the German G8 reform not only shortened academic secondary school (*Gymnasium*) from 9 to 8 years, but also increased the amount of weekly instructional time (by about 3.68 h or 12.5 % per week, see Hübner et al., 2022; Huebener et al., 2017), so that the total hours of instruction over the course of secondary school remained relatively constant. The G8 reform was one of the most fundamental reforms of the past decades and was accompanied by heated debates (Dörsam & Lauber, 2019).

Completing an academic-track school is still the primary way to fulfill the requirements for enrollment in a German university (Büchele, 2020). The *Gymnasium* is the highest of several secondary school tracks, which start after Grade 4 in most German federal states. Some states require a school track recommendation at the end of Grade 4 from elementary school teachers, while others leave the decision to students' parents. Students who start secondary school in the *Realschule* (intermediate-track school) or *Hauptschule* (lower-track school) and perform well have an opportunity to switch to an academic-track school after Grade 9 or 10, depending on the federal state (Biewen & Tapalaga, 2017). The main concern of parents, teachers, and researchers about the G8 reform was that increasing weekly instructional time could harm children's development (Kühn et al., 2013; Lehn, 2010; Marcus et al., 2020; Quis, 2018). Empirical evidence on the outcomes of the reform, however, varies across outcomes.

Andrietti and Su (2019), as well as Huebener et al. (2017), found that G8 students performed better than G9 students on the PISA tests in Grade 9; however, students with low socio-economic status benefitted less than students with a higher socio-economic status, or not at all. Additionally, Huebener and Marcus (2017) found that G8 students had a lower final grade point average (GPA) than G9 students, although their graduation rates were unchanged. Furthermore, as outlined by Hübner et al. (2017), G8 students exhibited lower performance in English reading and biology by the end of upper secondary school. Other studies suggested that the lower GPA of G8 compared to G9 students was likely driven by worse grades in mathematics (Büttner & Thomsen, 2015). Interestingly, Dahmann (2017) found no significant differences between G8 and G9 students in (fluid and crystalline) intelligence by the end of high school and Roth (2019) found no significant reform effects on social inequalities in upper secondary school attendance.

While Huebener and Marcus (2017) found that the increased grade repetition rates continued until the end of high school, Hübner et al. (2022) showed that the grade repetition rate was generally rather low at academic-track schools: There were extremely few repeaters in the double cohort, and there was only a very small difference between G8 and G9 students (in favor of G8 students). Additionally, Meyer et al. (2019) found only moderate changes in the incidence of grade repetition and dropout following the G8 reform, which they could not track back to the reform. Several studies showed that the G8 reform increased school-related stress and health problems (Hübner et al., 2017, 2022; Marcus et al., 2020; Quis, 2018) and reduced emotional stability among upper-secondary students (Dahmann & Anger, 2014). Additionally, Hofmann and Mühlenweg (2018) showed that the G8 reform significantly reduced adolescents' self-rated mental health status.

Doersam and Lauber (2019), Marcus and Zambre (2019), Meyer et al. (2019), as well as Meyer and Thomsen (2016, 2018) more closely investigated the influence of the reform on outcomes after school graduation. While Meyer et al. (2019) found only minor indications of reform-related selectivity regarding the decision to enroll into higher education, Marcus and Zambre (2019) found that G8 students were more likely to delay enrollment, drop out, or change their fields of study in higher education. The finding on delayed higher education entry is consistent with Meyer et al. (2019) and Meyer and Thomsen (2016). In contrast, Meyer and Thomsen (2018) and Doersam and Lauber (2019) did not find higher university dropout rates or differences in exam performance. However, the results of Meyer and Thomsen (2018) refer to only one German federal state and the results of Doersam and Lauber (2019) to only one university.

1.2. Present study

Building on the aforementioned studies, we examined the effects of the German G8 reform, which increased instructional time in lower secondary school while reducing the overall years of schooling by one year in German academic-track schools. We focused on reform-related changes in students' study time at university. More precisely, we examined time spent attending classes and on self-study. Moreover, we investigated whether students invest their potentially freed-up time in paid work during their studies. Finally, less concentration on one's studies could also manifest in later enrollment in higher education. Therefore, we additionally analyzed if the reform increased the number of months passing between high school graduation and university enrollment. This allowed us to investigate whether students were more likely to take a gap year. Though our literature review showed that examining instructional time has a long history, we could not identify any study investigating its potential long-term impact on study time after school.

From a life course perspective (e.g., Becker & Schulze, 2013), it is theoretically plausible that either processes of habituation (a) or of compensation (b) explain the effects of the examined policy reform. The habituation scenario (a) suggests that G8 students became used to a higher learning intensity in lower secondary school, which they maintained in higher education. If this was the case, we would expect to find that G8 students spent more hours attending classes and on self-study in higher education than G9 students. In contrast, the compensation scenario (b) suggests that the higher learning intensity in lower secondary school and associated reduced leisure time and health as well as increased stress created a desire among G8 students for more leisure time after school completion. If this was true, we would expect to find that G8 students spent less hours on attending classes and on self-study and took more time between high school completion and university entry than G9 students.

The aforementioned literature on the impact of the G8 reform on achievement found either no G8 reform effects or even a worse performance at the end of secondary school and university. Moreover, the cited health literature found negative effects of the reform. Therefore, the habituation scenario (a) seemed rather unlikely, and we expected to find evidence for the compensation scenario (b). Further support for the compensation scenario would be similar hours spent on paid work during the semester among G8 students and G9 students, as more time spent working during the semester would contradict the desire to experience more leisure time. G8 students had less free time during school, meaning that if scenario (b) was to hold true, students should not simply invest their freed-up time into more paid work. As additional support for scenario (b), G8 students could also take more time between graduation and university enrollment, as this would show that students were more likely to take a gap year after a more stressful time in secondary school. We intentionally did not look at student performance, because grades assigned in similar courses at different universities over multiple terms are not based on standardized tests and are thus not comparable across universities.

2. Methodology

The so-called G8 reform marked a pivotal educational transformation in several German states in the early 2000s (Homuth, 2017). To accelerate students' entry into the labor market without significantly altering their core curriculum, the reform redistributed the learning content of nine years into eight years. Importantly, the total number of hours of instruction required for graduation from academic-track schools remained constant, thus maintaining the time standards for student achievement. The reform primarily affected cohorts entering academic-track schools after the completion of primary school. Its implementation varied from state to state, with different ministries of education designing specific schedules in consultation with educational researchers and practitioners. Following the reform, the average number of instructional hours per week increased substantially, by about 3.68 h or 12.5 % (Hübner et al., 2022; Huebener & Marcus, 2017).

The start of the G8 reform differed between German federal states. The reform was implemented between 1999 and 2007 in all states but two. Rheinland-Pfalz never implemented the reform, and in Schleswig-Holstein, the reform was implemented one year after the last year covered by our dataset. Sachsen and Thüringen always had the G8 regime, as this was common in East Germany before reunification in 1990 and not changed afterwards. Therefore, we omitted these always-takers from the analysis. Similar to prior studies (e.g., Huebener et al., 2017), we also decided to drop Hessen, where the reform was implemented gradually over the course of three years and thus differed from the implementation in the other states. Due to the reduced number of years of schooling, the first G8 cohort graduated in the same year as the last G9 cohort. Hence, the first treatment cohort is a so-called "double cohort". Table 1 shows in which states and years students received their diploma under the G8 regime.

The reform implementation created a quasi-experimental setting, which can be exploited using a difference-in-differences (DiD) method. The DiD method aims at estimating a treatment effect by comparing the effect of a quasi-experimental treatment on a treatment group with a baseline measurement and to a corresponding measure for an untreated group. The resulting estimation of a causal effect is commonly referred to as the average treatment effect on the treated (ATT). The idea of DiD is to compare a treatment-induced change over time in the treatment group to a change without treatment in the control group. To calculate valid standard errors, DiD is typically estimated in a regression framework using two-way fixed effects (TWFE) models (e.g., Biewen & Schwerter, 2022; Hackenberger et al., 2021; Henry et al., 2020; Jacob et al., 2017; Zimmer et al., 2017). Such a regression model can be formalized as follows:

$$y_{ist} = \alpha + \lambda_s + \lambda_t + \rho\, Treatment_{st} + X_{ist}^{'}\beta + \varepsilon_{ist}$$

where index *i* is on the individual level, *s* is on an aggregated level like a state, and *t* stands for the time point. λ_s captures general state effects and λ_t captures time effects. The variable *Treatment* is equal to 1 for the treatment group after the treatment took place and 0 otherwise. The vector *X* includes all additional covariates. The classical DiD setting is the 2 × 2 setting, in which there are two periods and two states. In the second period, one state has received the treatment, and the unobservable counterfactual development is interpolated using the untreated state. To control for possible dependence within clusters (Bertrand et al., 2004), standard errors are usually clustered by the time dimension for each state. If the data are characterized by few (treated) clusters, the standard errors may still be biased toward zero (MacKinnon & Webb, 2017; Roth et al., 2023). To solve this issue, one can apply a wild cluster and subcluster bootstrap to obtain more reliable standard errors (Biewen & Schwerter, 2022; MacKinnon & Webb, 2017; Roth et al., 2023).

The German federal states implemented the G8 reform in different years (Table 1). In such cases, using TWFE to estimate the ATT can be troublesome (Callaway & Sant'Anna, 2021b; applied in Gándara & Li, 2020). First, TWFE assume a constant treatment effect. However, TWFE suffer from heterogeneity bias if the treatment effect is not homogenous. In this case, the parallel trend bias is positive, while the heterogeneity bias is negative. Ex-ante, it is unclear which of these biases dominates or whether they cancel each other out. Consequently, the direction of the estimated coefficient in TWFE can be incorrectly estimated due to these biases (Callaway & Sant'Anna, 2021b). Second, TWFE work well only in the classical 2 × 2 setting. If there are more time points with different treatment dates, TWFE aggregate all underlying 2 × 2 DiD treatment effects into one ATT. In doing so, TWFE put more weight on the underlying 2 × 2 DiD at the center of the time distribution and on those with more observations (Callaway & Sant'Anna, 2021b).

There are two methods to solve the TWFE treatment effect problem: weighted group-time ATT (Callaway & Sant'Anna, 2021b; Roth et al., 2023; Sun & Abraham, 2021) and imputation (Athey & Imbens, 2021; Borusyak et al., 2022). Because we observe only two never-treated states with relatively few observations, we must base our comparison on states that have not yet been treated. This is easier with weighted group-time ATT (Callaway & Sant'Anna, 2021b). Weighted group-time ATT are based on manually aggregated DiD models of each non-forbidden 2 × 2 DiD. Forbidden 2 × 2 DiD are those in which the already-treated states are used as a comparison group for the to-be-treated states. In other words, the main idea is to estimate the ATT of each 2 × 2 DiD and then aggregate the treatment effect of interest. This model works well even if (i) treatment effects are heterogeneous, as in the case of a staggered reform introduction, (ii) treatment effects change over time, (iii) short-run effects are more pronounced than long-run effects, and (iv) treatment effect dynamics differ when people are first treated in a recession relative to economic expansion years. In our study, (i) is the case as the reform was implemented in a staggered way, because educational decisions are taken at the state level in Germany. Moreover, (ii) changes in treatment effects over time are likely because the share of students affected by the reform changed as each additional German state adopted the reform, meaning that the treatment effect might have varied across states. Therefore, we rely on the R package *did* (Callaway & Sant'Anna, 2021a) and use not-yet-treated individuals as the control group. Given that we only have few clusters, we do not cluster by state, as the ordinary wild bootstrap performs better than the wild cluster or

Table 1
G8 Reform Introduction per State and High School Graduation Year.

	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015
Baden-Württemberg	–	–	–	–	–	–	+	+	+	+
Bayern	–	–	–	–	–	+	+	+	+	+
Berlin	–	–	–	–	–	–	+	+	+	+
Brandenburg	–	–	–	–	–	–	+	+	+	+
Bremen	–	–	–	–	+	+	+	+	+	+
Hamburg	–	–	–	–	–	–	+	+	+	+
Hessen	–	–	–	–	–	–	–	+	+	+
Mecklenburg-Vorpommern	–	–	+	+	+	+	+	+	+	+
Niedersachsen	–	–	–	–	+	+	+	+	+	+
Nordrhein-Westfalen	–	–	–	–	–	–	–	+	+	+
Rheinland-Pfalz	–	–	–	–	–	–	–	–	–	–
Saarland	–	–	–	+	+	+	+	+	+	+
Sachsen	+	+	+	+	+	+	+	+	+	+
Sachsen-Anhalt	–	+	+	+	+	+	+	+	+	+
Schleswig-Holstein	–	–	–	–	–	–	–	–	–	–
Thüringen	+	+	+	+	+	+	+	+	+	+

Note: – = Before implementation of G8 reform. + = After implementation of G8 reform. Rheinland-Pfalz never implemented the reform. The first G8 high school graduates in Schleswig-Holstein graduated in 2016, but we do not have data covering years after 2015.

subcluster bootstraps for few (treated) cluster (MacKinnon & Webb, 2017; 2018; Roodman et al., 2019). One feature of the ordinary wild bootstrap is that it uses randomly generated Rademacher weights (+1 or −1) with equal probabilities. The function *att_gt()* in the *R*-package *did* (Callaway & Sant'Anna, 2021a) uses the multiplier bootstrap by default, which basically amounts to perturbing the influence function by multiplying it by +1 or −1 weights. When including additional control variables, we use doubly robust estimations. Our tables and figures provide statistical significance tests and confidence intervals at the 5 % and 10 % significance levels in the figures.

We first present weighted group-time ATT without any additional variables. We also present event-study graphs to visualize the reform effect over time. Thereafter, we include control variables. Following Huebener et al. (2017), we thereby check threats to the identification strategy, i.e., changes in the composition of the students due to reform, by using the potential control variables as outcome variables. Moreover, we examine potentially heterogenous ATT by gender and social background, rerunning our analysis with and without control variables for the subsamples of men, women, first-generation students, and (at least) second-generation students. As additional robustness checks, we consider different coding scenarios for the double cohort, treating it as part of the treatment group initially (following Huebener et al., 2017), excluding it, and recoding the double cohort to a pre-treatment cohort (recoded DC) including a one-year anticipation period. Lastly, we compare the results of the weighted group-time ATT with a TWFE-regressions to show the difference between the new and old method. In the TWFE-regressions, we apply a *p*-value correction for few-treated clusters using the ordinary wild bootstrap (Fischer & Roodman, 2021; MacKinnon & Webb, 2017).

For the DiD method to produce valid estimates, we must ensure that the reform did not affect the group composition. Concretely, the reform must not encourage students to choose a *Gymnasium* (i.e., academic school track) or other schools. Likewise, families should not have moved to another state to evade the reform. Huebener et al. (2017), Huebener and Marcus (2017) and Anger and Dahmann (2014) have already critically examined and convincingly rejected these potential threats to the validity of our results. In fact, the mobility rates of citizens in Germany are generally low compared to those in other OECD countries (Kulu et al., 2021; Sánchez & Andrews, 2011). Most relocations in Germany happen within close ranges; moreover, parents' intentions to improve their children's education are rarely a motivation for relocations within Germany (Kemper, 2008). To additionally check the parallel trends assumption, we provide event-study graphs below.

3. Data

To exploit the described DiD setting, we used data from the German Student Social Survey (https://doi.org/10.21249/DZHW:ssypool:1.0.1). Since 1951, this survey provides nationally representative data on the socio-demographic characteristics, previous educational history, and current situation of students in German higher education. Using a simple random sampling procedure, the survey addresses cross-sections of students at most public and private higher education institutions every three to four years in the summer semester (for details, see Apolinarski et al., 2021). We analyzed data from the survey years 2003, 2006, 2009, 2012, and 2016. We used the first two waves mainly for pre-trend analyses, as no state had implemented the reform yet in 2003 and 2006. Even though the survey is conducted only every three to four years, each high school graduating year is represented because university students in different years of study are surveyed. Thus, the 2009 survey cohort includes not only first-year students who graduated in 2008, but also more advanced students who graduated before 2008.

As the reform only targeted academic-track schools, the secondary school type with the highest academic standards in Germany, we only considered students who graduated from this track. Furthermore, we dropped students who had been enrolled at university for more than 16 semesters, for whom the gap between high school graduation and survey year was larger than eight years, and students above the age of 30.

After applying these sample restrictions, our pooled dataset contained $N = 69{,}513$ students, of which 59 % were female (Table 2). Students in the sample had an average age of 23. The average number of semesters of university enrollment was 6.63. The vast majority (81 %) attended a university and about one-fifth (19 %) a university of applied sciences. Furthermore, 41 % of students had attended school after the implementation of the reform (treatment group). Only a few students were born outside Germany (3 %) and had children (2 %). Summary statistics for different areas of study and divided by G8 and G9 students are shown in the appendix (Tables A1, A2, and A3).

We investigated the effects of the G8 reform on eight outcome variables. These variables provide a comprehensive picture of students' study time investment: First, we examined time spent attending classes, time spent on self-study, and working hours during a typical week during the term. The question used to capture these items was "During the last 'typical' week of the semester, how many hours a day did you spend on the following activities? Indicate this for each day of the week in hours. Please round to full hours!". We aggregated the daily information into two distinct variables for workweeks (Monday-Friday) and weekends (Saturday-Sunday). We also investigated a dichotomous variable that assessed whether students worked for pay during the semester. Lastly, we looked at the number of months between high school completion and university entry.

Table 3 shows that students spent 12.41 h on average attending classes during the workweek and (as expected) hardly any time on the weekends. They further spent another 12.7 h on self-study per week – 9.2 h during the workweek and 3.5 h on the weekend. Moreover, they devoted about 3.6 h to paid work during the workweek and 1.2 h on the weekend. Around 64 % of students worked during the semester. The average number of months between high school completion and university entry was about 11 months.

Table 2
Sample Descriptives.

	Observations	Mean	SD	Min	Max
Treatment (1 = yes)	69,513	0.41	0.24	0	1
High school graduation year	69,513	2008.26	5.13	1998	2016
Female	69,513	0.59	0.49	0	1
Age	69,513	22.99	2.35	16	30
First-generation student	69,513	0.46	0.50	0	1
Semester of study	69,513	6.63	3.71	1	16
University (vs. university of applied sciences)	69,513	0.81	0.39	0	1
Child(ren)	69,513	0.02	0.12	0	1
Immigrant background	69,513	0.03	0.18	0	1

Note: SD = standard deviation, Min = minimum, and Max = maximum (of possible values of each variable). Descriptive statistics differentiated by G8 and G9 students and by area of study are presented in the appendix (Table A2).

Table 3

Descriptive Statistics of Outcome Variables.

	Observations	Mean	SD	Min	Max
Workweek					
Hours spent attending classes	69,513	12.41	11.49	0	71
Hours spent on self-study	69,513	9.20	10.11	0	80
Hours spent working	69,513	3.61	6.70	0	60
Weekends					
Hours spent attending classes	69,513	0.16	1.06	0	24
Hours spent on self-study	69,513	3.48	4.48	0	38
Hours spent working	69,513	1.22	3.18	0	32
Non-weekly variables					
Working during the semester	67,906	0.64	0.48	0	1
Months between high school completion and university entry	69,513	11.26	12.34	0	106

Note: SD = standard deviation, Min = minimum, and Max = maximum (of possible values of each variable). Descriptive statistics differentiated by G9 and G8 students and by area of study are presented in the appendix (Table A3).

Table 4

Treatment Effects of the G8 Reform on Study Time at University.

	ATT	SE	Lower 95% CI	Upper 95% CI
Workweek				
Hours spent attending classes	−1.745*	0.480	−3.002	−0.488
Hours spent on self-study	−1.401*	0.371	−2.384	−0.418
Hours spent working	−0.062	0.197	−0.568	0.444
Weekends				
Hours spent attending classes	−0.003	0.029	−0.070	0.064
Hours spent on self-study	−0.519*	0.180	−1.011	−0.028
Hours spent working	−0.165	0.121	−0.474	0.144
Non-weekly variables				
Working during the semester	0.043	0.021	−0.014	0.099
Months between high school completion and university entry	1.792*	0.349	0.905	2.680

Note: ATT refers to the weighted group-time average treatment effect on the treated (Callaway & Sant'Anna, 2021b). Standard errors (SE), lower and upper 95 % confidence intervals (CI) are adjusted for multiple testing using ordinary bootstrapped standard errors. $N = 69{,}513$. $^*p < 0.05$.

4. Results

Our baseline results indicate significant G8 treatment effects on time spent attending classes during the workweek, hours spent on self-study during the week and on the weekend, and on the time gap between high school completion and university entry (Table 4). Students affected by the G8 reform spent about 1.7 h less attending classes during the week. In addition, G8 students spent about 1.4 h less on self-study during the workweek and 0.5 h less on the weekend. We also found an average increase of 2 months in the duration between high school graduation and university entry. There was no statistically significant effect on time spent attending classes during the weekend, hours spent on paid work (during the workweek or the weekend), and the general probability of working. Overall, these results align with the compensation scenario (b), which suggests that G8 students spend less time on self-study at university, without increasing their working time.

We further decomposed the general ATT using an event-study design (Fig. 1) in which the treatment effects in all states were aggregated in terms of normalized time relative to the reform implementation. In so doing, we tested (a) whether there was a pre-trend that violates the common trends assumption and (b) whether the ATT were constant across all post-treatment periods. This elucidates whether (a) the parallel trends assumption holds and whether (b) the treatment effects changed over time.

First, we did not observe any statistically significant differences in the pre-treatment periods over time, supporting the validity of the common trend assumption concerning all examined outcome variables. For time spent attending classes during the workweek, the effect was negative but not statistically significant when the treatment was first introduced. For the following two post-treatment periods, we found negative and statistically significant effects. Thereafter, the estimates were constant from the third throughout the eighth post-treatment periods, but the confidence intervals widened, resulting in statistically insignificant effects. Notably, G8 students graduated from high school between 2007 and 2013, and the last data collection point was in 2015. Thus, our data included only a few students from some larger states with more than three post-treatment periods. This explains the less efficient estimations for the later periods.

The effect of the G8 reform on hours spent on self-study during the workweek was significant for the double cohort and the subsequent two graduation years. Again, the estimation results were relatively stable but less efficient for the later post-reform years. For hours spent on self-study on the weekend, we found a trend of decreasing ATT over time up to the fifth post-treatment period. Still, only the effect for the second post-treatment period was statistically significant. Lastly, for the months between high school graduation and university entry, the effects were significant for the first to fifth post-treatment periods. The graphs for time spent in class on the weekend and all three work-related outcomes did not reveal any additional information compared to Table 4, as there were no significant differences.

To check for differences between federal states, we included control variables for gender, age, and whether students attended a university or a university of applied sciences. Because the reform affected students beginning in Grade 5, any student characteristics measured after Grade 5 might represent intermediate outcomes or "bad" controls (Angrist & Pischke, 2009). Consequently, a reduction in the treatment effects after adding the additional control variables would not necessarily imply that the treatment was less effective, but rather that the treatment effects were partly explained by the intermediate outcomes. For ease of interpretation, we compared the results presented in Table 4 graphically with the estimation results using additional control variables in Fig. 2. We found no change in the interpretation of whether the treatment effects were statistically unequal to zero at the 10 % significance level. Only for hours spent on self-study during the weekend does the 5 % interval now overlap zero. Thus, our previous estimation results were robust and cannot be explained by the additional control variables.

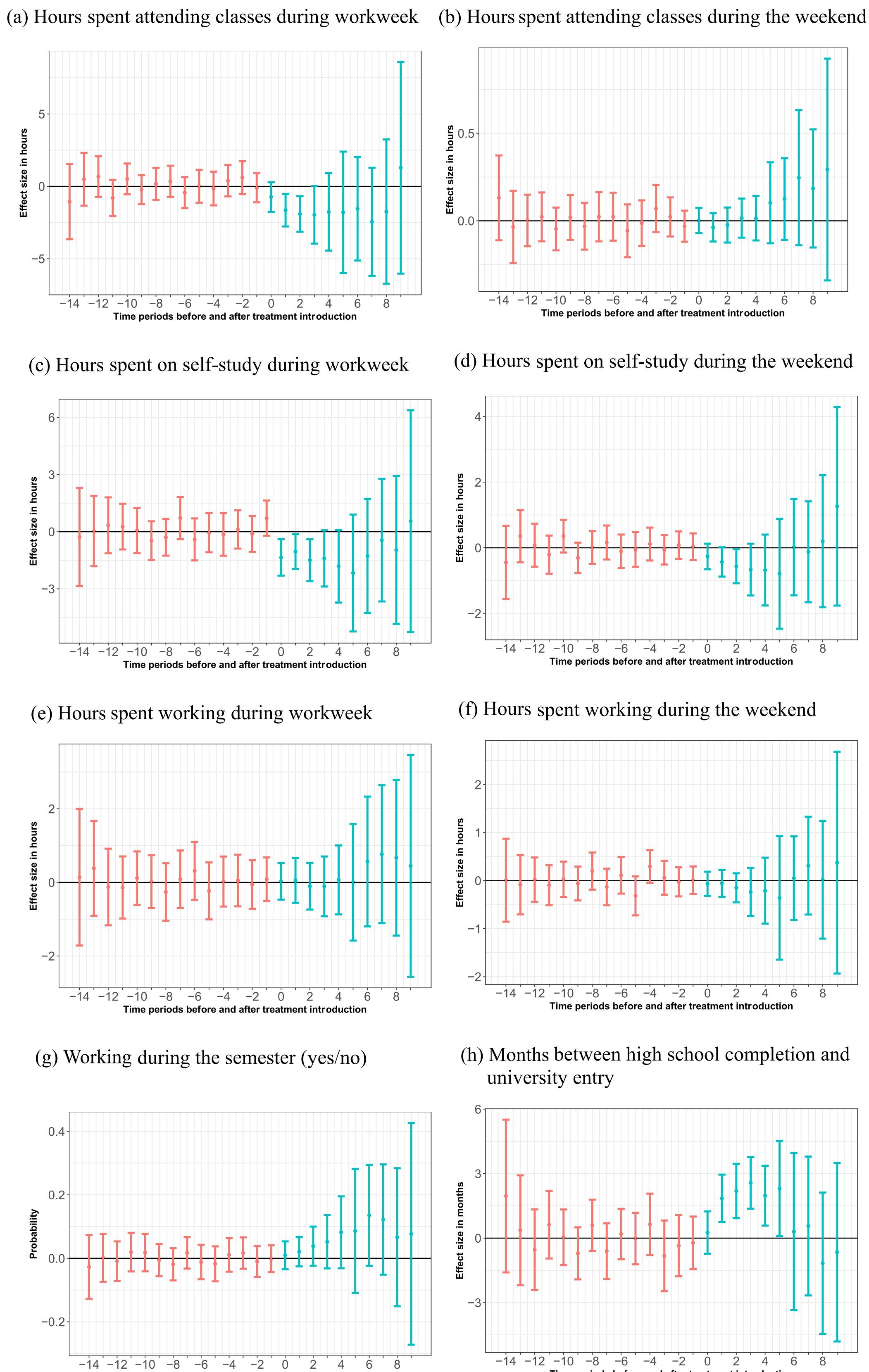

Fig. 1. Event-Study Results.
Note: The x-axis shows the time in years relative to the treatment. The red lines (values on the x-axis less than zero) indicate years before the treatment, turquoise lines (values on the x-axis greater than or equal to zero) indicate years after the treatment. The y-axis shows the effect of the treatment on the respective outcomes. 95 % confidence intervals were adjusted for multiple testing using ordinary bootstrapped standard errors.

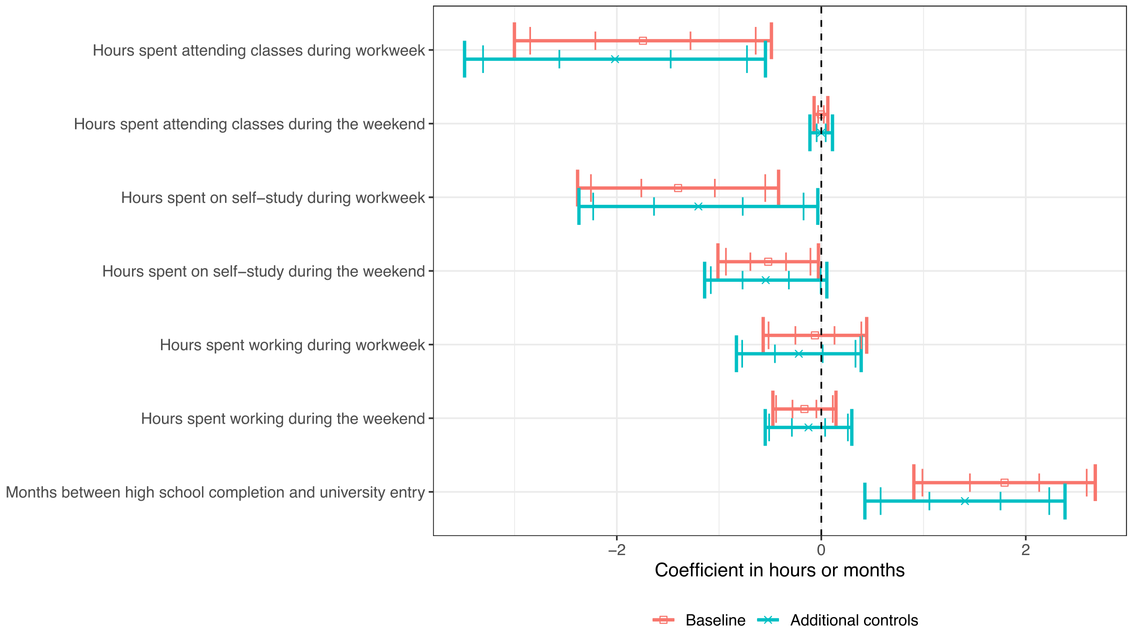

Fig. 2. Comparing Baseline Results with Estimations Including Additional Control Variables.
Note: We included the 95 % and 83.4 % confidence intervals for each estimate. We provided the 95 % confidence intervals to highlight whether the estimates were statistically different from zero at the 5 % level. We also included the 83.4 % confidence intervals to highlight whether the point estimates from the base regression (red, square) and the regression with additional control variables (turquoise, cross) were significantly different from zero at the 5 % level (Knol et al., 2011). We did not report estimates for working during the semester because the estimated reform effects was not significant and because the variable was not captured in hours or months but coded as a dichotomous variable.

5. Sensitivity analyses

5.1. Heterogenous reform effects by gender and social background

Following the G8 literature (Huebener et al., 2017), we explored potential heterogeneity in the treatment effects between first- and second-generation students (as an operationalization of social background) as well as between women and men (Fig. 3). An analysis for the subgroups of men vs. women revealed that the ATT on spending time attending classes during the workweek as well as on self-study during the workweek and on the weekend were partly driven by the estimates for men. Interestingly, whereas the ATT for men were statistically significant different from zero for all three aforementioned outcomes (spending time attending classes during the workweek as well as self-study during the workweek and on the weekend), this was not the case for women. Notably, although the ATT for hours spent working during the weekend were statistically insignificant for both men and women, the positive ATT for men was significantly different from the negative ATT of women, indicating that men were more likely to work longer hours than women if they work.

Moreover, we found that the ATT for spending time attending classes during the workweek was driven by the estimates for second-generation students. However, for time spent on self-study, the ATT for both second and first-generation students were not significantly different from zero. Our results suggest that treatment heterogeneity was more pronounced between men and women than between second and first-generation students. Overall, however, we found that the effects of the G8 reform on study time in higher education were rather similar across students of different gender and social background.

5.2. Double cohort

As the first treatment period also contained students from the G9 regime, we re-estimated our models considering the double cohort a pre-treatment cohort (recoded DC) including a one-year anticipation period. Additionally, we estimated models in which we completely dropped the double cohort (no DC). Again, all results were relatively robust (Fig. 4). Only for self-study during the week did the ATT shift clearly toward zero when assigning the double cohort to the non-treatment period without an anticipation period. This can be easily explained by the fact that in Fig. 1c, the double cohort has a significantly negative ATT. Additionally, not including the double cohort led to generally higher ATT, except with the months between high school completion and university entry. Other than that, the results were not affected much by the coding of the double cohort. Therefore, our main conclusions are unaffected by how we treat the double cohort.

5.3. Composition of treatment and control groups

To ensure the robustness of our DiD analyses, we examined possible compositional changes in the cohorts affected by the reform. Specifically, we tested whether the reform affected the selection of students into higher education by examining students' personal characteristics, choice of study area, and type of degree. Significant changes in the composition of student bodies by personal characteristics, study areas, or degree types might suggest that the reform changed the patters of selection into higher education and thus students' average study time. Such a scenario could explain our observed treatment effects.

Supporting our results, however, we did not find any significant reform-related compositional changes regarding students' personal

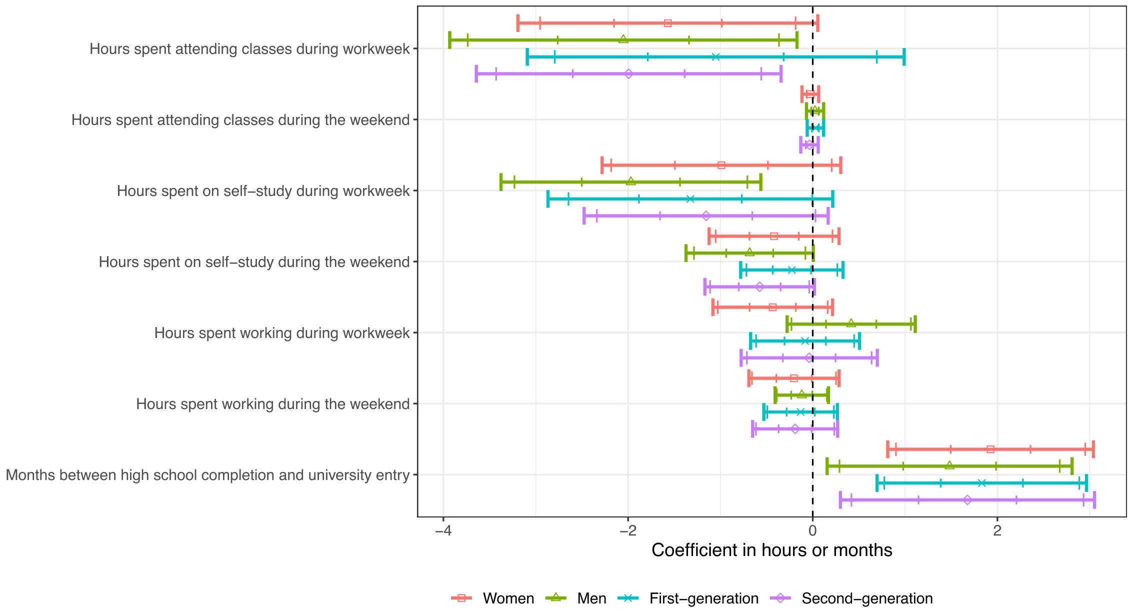

Fig. 3. Comparing the Reform Effects across Subgroups.
Note: We included the 95 % and 83.4 % confidence intervals for each estimate. We provided the 95 % confidence intervals to highlight whether the estimates were statistically different from zero at the 5 % level. We also included the 83.4 % confidence intervals to highlight whether the point estimates for the subgroups were significantly different from zero at the 5 % level (Knol et al., 2011).

characteristics, study areas, and types of degrees (Tables 5 and 6). Only students' age differed significantly between pre- and post-reform cohorts at the 95 % level; moreover, the share of first-generation students differed significantly at the 90 % level (Table 5). The first difference, however, is by construction: As the reform was conducted more recently in the bigger German federal states, there were not yet so many older post-reform students enrolled at university. Importantly, younger students usually invest more time in their studies than older students (Hauschildt et al., 2015, 2018; Pellizzari & Billari, 2012). Similarly, first-generation students tend to be less engaged in their studies and perform worse than second-generation students (Canning et al., 2020). Thus, having more younger students and fewer first-generation students in the G8 sample might (if at all) have biased our estimates toward zero, leading (at worst) to lower bound estimates of the true treatment effects.

Besides the age difference, we did not find any significant differences between pre- and post-reform cohorts. Consequently, we are highly confident that the reform did not unintentionally influence students' interest in different study areas or types of degrees – and thus that we actually measured the effects of G8-related changes in instructional time.

5.4. Comparing weighted group-time with TWFE estimates

This section demonstrates the necessity of using the DiD method introduced by Callaway and Sant'Anna (2021b). Typically, multiple specifications with different sets of control variables are included sequentially when using the TWFE. However, the biggest problem are *forbidden DiDs* in TWFE estimations, which compare always-takers or already-takers (students from states that already implemented the reform) with to-be-treated students. The following graphs show the TWFE estimations for five different models. Following Huebener et al. (2017), the first TWFE model includes the general DiD variables and a dummy for the double cohort (Model 1). The second model includes the control variables for gender, age, semester of study, enrollment in a university vs. university of applied sciences, having at least one child, nationality and first-generation status (Model 2). The third model added the type of degree students were pursuing (e.g., bachelor's or master's degree) and a dummy variable for each study area (Model 3). Table 7 presents the respective estimation results with ordinary wild bootstrap *p*-values.

The TWFE yielded estimates biased heavily toward zero. For example, instead of an estimated reduction of almost two hours, the TWFE estimated treatment effects of about half an hour up to one hour for hours spent attending classes during the workweek. This estimation highlights the problem with TWFE, as it underestimated the G8 reform effects. Only for the number of months between high school completion and university were the estimation results similar with both methods, albeit only once we included additional control variables. In terms of significance, however, the same estimates were significant, as in Table 4.

6. Discussion and conclusion

Adopting a life course perspective, we examined whether changes in the learning environment in the early life course have long-lasting effects on students' study time later in life. We investigated the effects of the G8 reform in Germany, which reduced secondary school by one year and increased average instructional time by about 3.68 h or 12.5 % per week. Instructional time in the remaining eight years of secondary school was increased to compensate for the omitted school year, so that a similar amount of content could be covered. We tested whether this quasi-experimental setting influenced the duration between high school completion and university entry, students' investment in different forms of study time, and the likelihood of working during the semester.

Our results showed that G8 students spent less time attending classes at university during the workweek than G9 students. The G8 reform also reduced hours spent on self-study during the workweek and weekends. Thus, compressing students' time in school while increasing average weekly instructional time had detrimental effects on study time at university. Furthermore, G8 students did not use their freed-up time to work more and tended to enroll at university after longer time gaps than G9 students.

We started from two theoretically plausible scenarios explaining possible G8 reform effects, that is, a habituation scenario (a) and a

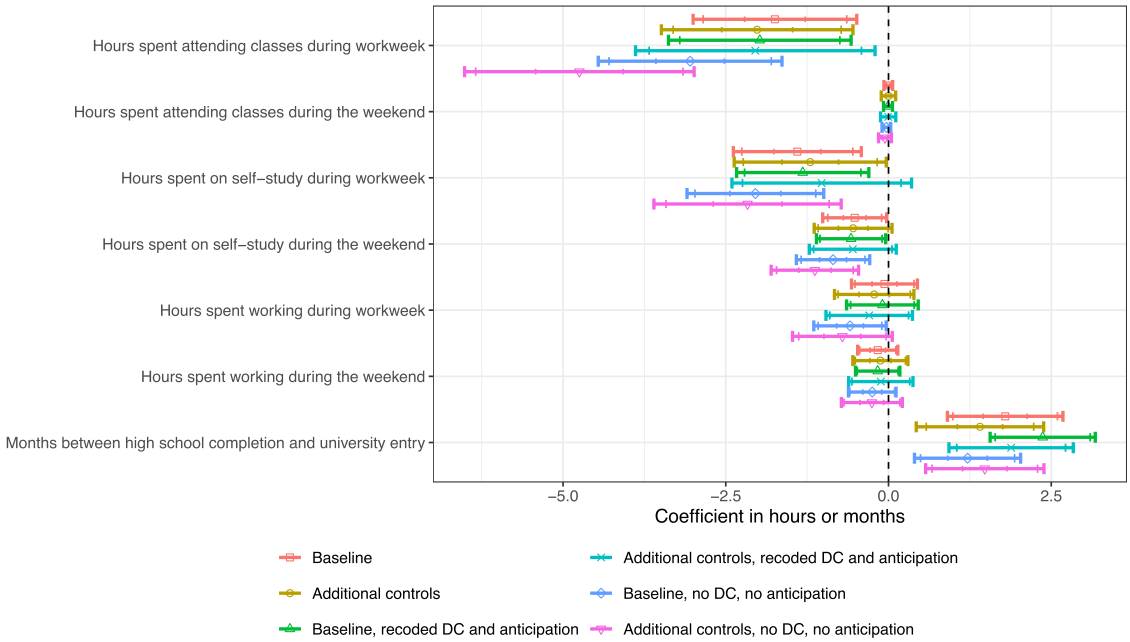

Fig. 4. Robustness Check Including Various Treatments of the Double Cohort.
Note: As in Fig. 2, we included the 95 % and 83.4 % confidence intervals for each estimate. In addition to the baseline regressions (red, square) and regressions including additional control variables (yellow, triangle) presented above, we added regressions in which the double cohort was recoded as a pre-treatment cohort without anticipation (recoded DC, no anticipation) and with anticipation (recoded DC and anticipation) – each of these treatments of the double cohort once without additional control variables (green, cross; purple, triangle pointing downwards), and once with additional control variables (blue, diamond). 95 % confidence intervals were adjusted for multiple testing using ordinary bootstrapped standard errors.

Table 5
Treatment Effects of the G8 Reform on Students' Characteristics.

Variable	Observations	ATT	SE	95 % CI Lower CI	95 % CI Upper CI	90 % CI Lower CI	90 % CI Upper CI
Female	69,513	0.017	0.022	−0.042	0.075	−0.034	0.068
Age	69,513	−0.415*	0.077	−0.625	−0.205	−0.591	−0.239
Number of children	69,513	−0.014	0.007	−0.030	0.002	−0.030	0.003
Foreign	69,513	−0.017	0.012	−0.046	0.011	−0.045	0.010
First-generation student	69,513	−0.047^{+}	0.019	−0.097	0.002	−0.091	−0.003
Married (vs. single or in a relationship)	67,093	−0.007	0.004	−0.017	0.003	−0.164	0.003
Single (vs. married or in a relationship)	67,093	0.014	0.017	−0.030	0.058	−0.023	0.050
Father with university entrance qualification	65,389	0.020	0.016	−0.023	0.063	−0.020	0.061
Mother with university entrance qualification	66,312	−0.004	0.017	−0.050	0.042	−0.044	0.036
Father without school-leaving certificate	65,389	0.000	0.003	−0.008	0.009	−0.007	0.007
Mother without school-leaving certificate	66,312	0.003	0.003	−0.003	0.010	−0.002	0.009

Note: ATT refers to the weighted group-time average treatment effect on the treated (Callaway & Sant'Anna, 2021b). The last six variables were not included as control variables in any analysis to prevent a reduction of observations. Married: Dummy variable indicating whether students are married (vs. single or in a relationship). Single: Dummy variable indicating whether students are single (vs. married or in a relationship). With university entrance qualification: 1 = university entrance qualification [Abitur], 0 = secondary school qualification [Realschulabschluss], lower secondary school qualification [Hauptschulabschluss], or no school-leaving certificate. With no school-leaving certificate: 1 = no school-leaving certificate, 0 = university entrance qualification, secondary school qualification, or lower secondary school qualification. Standard errors (SE), lower and upper confidence intervals (CI) were adjusted for multiple testing using ordinary bootstrapped standard errors. * $p < 0.05$, $^{+} < 0.10$.

compensation scenario (b). Our empirical results align with the compensation scenario (b) in that G8 students spent less time studying without working more and enrolling at university after longer time gaps. Thus, G8 students opted for more leisure time after school completion than G9 students. The estimated ATT were particularly pronounced in the first periods after the introduction of the treatment. Thereafter, the treatment effects were reasonably stable, but not statistically significant due to large standard errors. One possible explanation for the lack of significance in the later post-reform periods is the relatively small number of G8 students from states that implemented the reform in 2012 or 2013. For example, only three cohorts of students were affected by the reform in those states in which the first G8 students graduated in 2013.

Our results are consistent with previous findings on the G8 reform suggesting rather negative effects of reform exposure on educational attainment: Several studies found that G8 students delayed their university entry, changed their fields of study more frequently, and were more likely to drop out of higher education (Marcus & Zambre, 2019; Meyer et al., 2019; Meyer & Thomsen, 2018). These patterns align with

Table 6
Treatment Effects of the G8 Reform on Students' Study Area and Degree Type.

Panel A				95 % CI		90 % CI	
Study Area	Observations[S]	ATT	SE	Lower CI	Upper CI	Lower CI	Upper CI
Agricultural, forestry, nutritional sciences	1,722	0.016	0.015	−0.022	0.056	−0.018	0.051
Medicine	5,227	0.002	0.003	−0.006	0.010	−0.005	0.010
Engineering	13,012	−0.028	0.020	−0.081	0.026	−0.075	0.020
Art	2,138	0.003	0.015	−0.037	0.042	−0.032	0.037
Mathematics & sciences	12,090	0.007	0.009	−0.016	0.030	−0.014	0.028
Law, business, economics, and social scienes	20,486	−0.001	0.001	−0.003	0.001	−0.003	0.001
Sport	945	0.003	0.005	−0.010	0.016	−0.008	0.015
Languages	14,662	−0.004	0.019	−0.054	0.044	−0.048	0.039
Panel B				95 % CI		90 % CI	
Degree Type	Observations[D]	ATT	SE	Lower CI	Upper CI	Lower CI	Upper CI
Other degree	272	−0.029	0.019	−0.079	0.021	−0.073	0.015
Bachelor (without teaching degree)	27,146	0.011	0.011	−0.016	0.037	−0.013	0.035
Bachelor (teaching degree)	2,823	0.008	0.013	−0.026	0.042	−0.021	0.038
Art college degree	9,251	0.007	0.005	−0.005	0.018	−0.005	0.018
University vs. university of applied sciences	2,956	−0.003	0.001	−0.006	0.000	−0.006	0.000
Not striving for a degree	153	0.005	0.004	−0.003	0.014	−0.003	0.014
Magister	2,723	0.000	0.001	−0.003	0.003	−0.003	0.003
Master (without teaching degree)	8,271	0.003	0.012	−0.029	0.034	−0.024	0.030
Master (teaching degree)	1,120	−0.008	0.011	−0.035	0.018	−0.033	0.016
Doctorate degree	623	0.000	0.000	−0.001	0.001	−0.001	0.001
State examination (without teaching degree)	8,582	0.004	0.002	−0.000	0.009	−0.000	0.009
State examination (teaching degree)	6,576	0.002	0.002	−0.001	0.006	−0.001	0.006

Note: Observations[S] shows how many students major in the respective study area. Observations[D] indicates how many students major in the respective degree type. ATT refers to the weighted group-time average treatment effect on the treated (Callaway & Sant'Anna, 2021b). Standard errors (SE), lower and upper confidence intervals (CI) were adjusted for multiple testing using bootstrapped standard errors. $N = 69{,}513$. * $p < 0.05$, $^{+} < 0.10$.

Table 7
Comparing Weighted Group-Time and TWFE Estimates.

	Two-Way Fixed-Effects Regressions		
	Model 1	Model 2	Model 3
Workweek			
Hours spent attending classes	−0.511*	−0.449*	−0.403*
	[0.002]	[0.004]	[0.007]
Hours spent on self-study	−0.541*	−0.589*	−0.546*
	[0.001]	[0.000]	[0.002]
Hours spent working	0.008	0.002	−0.005
	[0.923]	[0.982]	[0.961]
Weekends			
Hours spent attending classes	−0.012	−0.011	−0.010
	[0.463]	[0.480]	[0.528]
Hours spent on self-study	−0.181*	−0.186*	−0.173*
	[0.008]	[0.004]	[0.011]
Hours spent working	0.025	0.047	0.042
	[0.625]	[0.309]	[0.392]
Non-weekly variables			
Working during the semester	0.008	0.008	0.004
	[0.323]	[0.289]	[0.597]
Months between high school completion and university entry	0.716*	1.828*	1.830*
	[0.002]	[0.000]	[0.000]

Note: The ordinary wild bootstrap (Fischer & Roodman, 2021) *p*-values for the TWFE models are shown in rectangular brackets. Model 1 included no additional control but a dummy for the double cohort, Model 2 included students' gender, age, semester, first-generation status, type of university, and nationality. Model 3 additionally included degree type and area of study fixed effects. TWFE results using normal clustered SE are shown in the manuscript in Fig. A1. * $p < 0.05$, + $p < 0.10$.

our finding that G8 students are less focused on higher education attainment. This finding underscores a new facet of a broadly visible pattern of the G8 reform having possibly detrimental effects. The homogeneous treatment effects on students' from different social backgrounds also align with previous findings (Büttner & Thomsen, 2015; Huebener et al., 2017; Huebener & Marcus, 2017). Additionally, the tendency for a more negatively pronounced effect on men aligns with results by Huebener et al. (2017), who found that boys were more negatively affected by the reform than girls.

While G8-induced increases in average performance were visible among 9th graders in PISA data (Andrietti & Su, 2019; Huebener et al., 2017), negative effects on grade repetition rates, graduation rates, and GPA were found at the end of high school (Huebener & Marcus, 2017). A possible explanation for why G8 students are more likely to drop out of higher education (Marcus & Zambre, 2019) despite similar levels of intelligence, a similar personality development, unchanged levels of social inequality at the end of high school, and the absence of compelling evidence for different achievement levels of G8 students at university (Dahmann, 2017; Dörsam & Lauber, 2019; Meyer & Thomsen, 2018; Roth, 2019; Thiel et al., 2014) is the lower study time of G8 students we found in our analysis. Our results are of particular importance when reconsidering the ample evidence showing that more effort at university (e.g., preparing for classes, attending classes, practicing after classes, and especially self-study) leads to better educational performance (Andrietti & Velasco, 2015; Arulampalam et al., 2012; Bonesrønning & Opstad, 2015; Bratti & Staffolani, 2013; Ersoy, 2021; Grave, 2011; Metcalfe et al., 2019; Schwerter et al., 2022a, 2022b; Stinebrickner & Stinebrickner, 2008). Less effort on the part of G8 students might result from long-term negative effects of increased stress and reduced health at high school: Students who experience fatigue, feel overwhelmed, experience stress, and have overall poorer health (especially at the end of upper secondary school; Hübner et al., 2017, 2022; Marcus et al., 2020; Quis, 2018) might reduce their study time in higher education, when they have the autonomy to do so. They might do so to compensate, at least to a certain degree, for more stressful school years. In summary, the existing empirical evidence tends to confirm the main concern of parents, teachers, and researchers that more intensive daily instruction can have adverse effects on children's development in the long run (Kühn et al., 2013; Lehn, 2010).

The differential effects of the G8 reform on men and women might be explained by women's generally higher conscientiousness (Verbree et al., 2023), greater self-discipline (Duckworth et al., 2019; Duckworth & Seligman, 2006), and better grades (Blanden, 2020; Duckworth et al., 2019; Friedman-Sokuler & Justman, 2016; Lubinski & Benbow, 2006). Women's higher levels of conscientiousness, self-discipline, and academic performance could make them less susceptible to the effects of the reform, as they might be inherently more motivated or better prepared to navigate academic challenges induced by systemic changes. Notably, a formal test of the relevance of personality traits for gender disparities in the context of the G8 reform needs to be conducted in future studies that have adequately captured measures for these traits.

Our results are robust to several sensitivity checks and align with previous results. Still, our study has limitations, which also highlight ways forward for future research on the long-term effects of educational reforms. To begin with, the German Student Social Survey assessed students only every three to four years. As G8 students seem to be more likely to drop out of higher education (Marcus & Zambre, 2019), the remaining students may be particularly determined to complete their studies. However, as reported in previous studies, students who drop out are less likely to attend class and invest time in self-study (Bernardo et al., 2016). These patterns suggest that their higher likelihood of dropout likely leads to an underestimation of the ATT of the G8 reform. Considering this fact and that our G8 sample contains more younger students and less first-generation students, our estimations are likely conservative.

Our data on the number of hours spent studying were reported subjectively. While minor differences to actual hours spent studying are thus theoretically possible, it is unlikely that the reform changed students' response behavior in this respect, and therefore unlikely that the subjective measurement of our dependent variables produced notably biased ATT of the G8 reform. Furthermore, our data did not allow us to determine whether students only invested less time or whether they also obtained fewer ECTS credits per semester. Thus, we could not determine whether G8 students made less formal progress in their studies and graduated later from university than G9 students. We are not aware of large-scale data allowing researchers to test this.

We observed little change in the composition of student bodies due to the reform. Where we observed small changes, these were likely to result in lower-bound estimates. Other studies reported some selection into university as a result of the reform (Marcus & Zambre, 2019; Meyer et al., 2019; Meyer & Thomsen, 2016). As our data only included students in higher education, we could not further examine possible selection patterns in this study. Similarly, we were not aware of data allowing us to further examine what and how students learn in their time devoted to studying. Considering that G8 students do not only spend less time on average on class attendance and self-study in higher education, but also tend to receive similar or worse grades (Huebener & Marcus, 2017), exhibit a higher propensity to drop out of higher education (Marcus & Zambre, 2019), and report school-related stress and health problems more frequently (Quis, 2018), we have no reason to assume that G8 students learned to invest their study time more efficiently. Yet, future research could properly test this hypothesis. Analyzing study efficiency in addition to study time would provide further valuable evidence on whether the G8 reform induced a habituation or a compensation scenario.

Future research could also more systematically examine the implementation of the G8 reform at the school level. The reform foresaw a uniform framework for lower and upper secondary students within the different German states. Furthermore, the reform required schools to implement changes in the majority or all subjects and did not lead to a cut in the number of subjects. However, our data did not allow us to properly examine how the reform was implemented within individual schools. For instance, it would have been highly interesting to know whether study time was equally condensed in all subjects. The assumption of a uniform implementation of the reform is supported by the fact that schools within the same state typically use similar textbooks, indicating relatively homogenous curriculum compressions.

From a life-course perspective, it would also be important to examine whether the lower study time of G8 students in higher education extends to later periods in the life course. Are G8 students also less likely to enroll in professional developmental training later in life? This could be concerning in rapidly changing and increasingly digitized labor markets (Acemoglu et al., 2020, 2021; Acemoglu & Restrepo, 2019) – in which lifelong learning will be important for meeting future labor market demands (OECD, 2019).

To our knowledge, no prior study had investigated the extent to which policy reforms implemented in lower secondary school created path dependencies that influence study time in the life course after school completion. We addressed this research gap by showing that the German G8 school reform, which increased the weekly instructional time in lower secondary school for academic-track students, decreased students' time on class attendance and self-study in higher education. Additionally, we found no statistically significant reform effects on students' probability to work but that G8 students delayed their enrollment in higher education. Our study also advanced the corresponding methodological literature by presenting weighted group-time ATT (Callaway & Sant'Anna, 2021b), which overcome several problems of the TWFE in DiD frameworks. Our results suggest that researchers and policymakers should more seriously consider potential long-term effects when developing, implementing, and evaluating school reforms.

Declaration of generative AI and AI-assisted technologies in the writing process

During the preparation of this work, the authors used *DeepL* for language edits. After using this tool, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication.

Declaration of competing interest

We declare that we have no conflicts of interest.

Data availability

The authors do not have permission to share the analyzed data. However, these data are available via the DZHW Research Centre (https://doi.org/10.21249/DZHW:ssypool:1.0.1).

Acknowledgment

Jakob Schwerter is financed by the project "From Prediction to Agile Interventions in the Social Sciences (FAIR)", which is receiving funding from the program "Profilbildung 2020", an initiative of the Ministry of Culture and Science of the State of Nordrhein-Westfalen. The sole responsibility for the contents of this publication lies with the authors.

Appendix

Table A1
Descriptive Results for the Five Most Frequent Study Areas in Germany.

	Health Sciences		Engineering		Natural Sciences		Social Sciences		Languages and Cultural Studies	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Workweek										
Hours spent attending classes	16.68	14.08	10.96	11.89	14.57	13.09	11.09	10.51	12.25	9.18
Hours spent on self-study	9.81	10.65	7.90	10.33	10.41	10.73	8.68	9.93	9.98	9.25
Hours spent working	2.06	5.28	3.33	7.08	3.23	6.48	3.84	7.01	4.43	6.53
Weekend										
Hours spent attending classes	0.19	1.31	0.11	0.95	0.13	0.99	0.18	1.07	0.17	1.07
Hours spent on self-study	4.57	5.47	2.83	4.29	4.00	4.71	3.22	4.29	3.65	4.23
Hours spent working	1.40	3.76	0.81	2.62	1.08	2.92	1.18	3.12	1.55	3.46
Non-weekly variables										
Working during the semester	0.54	0.50	0.61	0.49	0.59	0.49	0.65	0.48	0.71	0.45
Months between high school completion and university entry	13.46	16.86	10.86	11.75	9.49	9.90	12.40	13.38	10.12	10.37

Note: SD = standard deviation.

Table A2
Sample Descriptives Differentiated by G9 and G8 Students.

	G9					G8				
	N	Mean	SD	Min	Max	N	Mean	SD	Min	Max
Double cohort	46,763	0.00	0.00	0	0	23,733	0.22	0.42	0	1
High school graduation year	46,763	2005.69	4.30	1998	2016	23,733	2013.33	1.80	2007	2016
Female	46,763	0.59	0.49	0	1	23,733	0.60	0.49	0	1
Age	46,763	23.58	2.26	16	30	23,733	21.83	2.10	17	30
First-generation student	46,763	0.46	0.50	0	1	23,733	0.47	0.50	0	1
Semester of study	46,763	7.24	3.89	1	16	23,733	5.44	2.99	1	16
University (vs. university of applied sciences)	46,763	0.83	0.37	0	1	23,733	0.76	0.43	0	1
Child(ren)	46,763	0.02	0.14	0	1	23,733	0.01	0.09	0	1
Immigrant background	46,763	0.03	0.17	0	1	23,733	0.04	0.20	0	1

Note: SD = standard deviation, Min = minimum, and Max = maximum (of possible values of each variable).

Table A3
Description of the Outcome Variables Differentiated by G9 and G8 Students.

	G9					G8				
	N	Mean	SD	Min	Max	N	Mean	SD	Min	Max
Workweek										
Hours spent attending classes	46,763	14.61	11.25	0	71	23,733	8.16	10.71	0	58
Hours spent on self-study	46,763	11.07	10.35	0	80	23,733	5.61	8.51	0	60
Hours spent working	46,763	4.42	7.24	0	60	23,733	2.06	5.20	0	60
Weekend										
Hours spent attending classes	46,763	0.19	1.19	0	24	23,733	0.09	0.76	0	20
Hours spent on self-study	46,763	4.12	4.64	0	38	23,733	2.25	3.88	0	25
Hours spent working	46,763	1.46	3.45	0	32	23,733	0.75	2.52	0	24
Non-weekly variables										
Working during the semester	44,524	0.66	0.47	0	1	23,382	0.61	0.49	0	1
Months between high school completion and university entry	46,763	12.35	13.37	0	106	23,733	9.14	9.64	0	100

Note: SD = standard deviation, Min = minimum, and Max = maximum (of possible values of each variable).

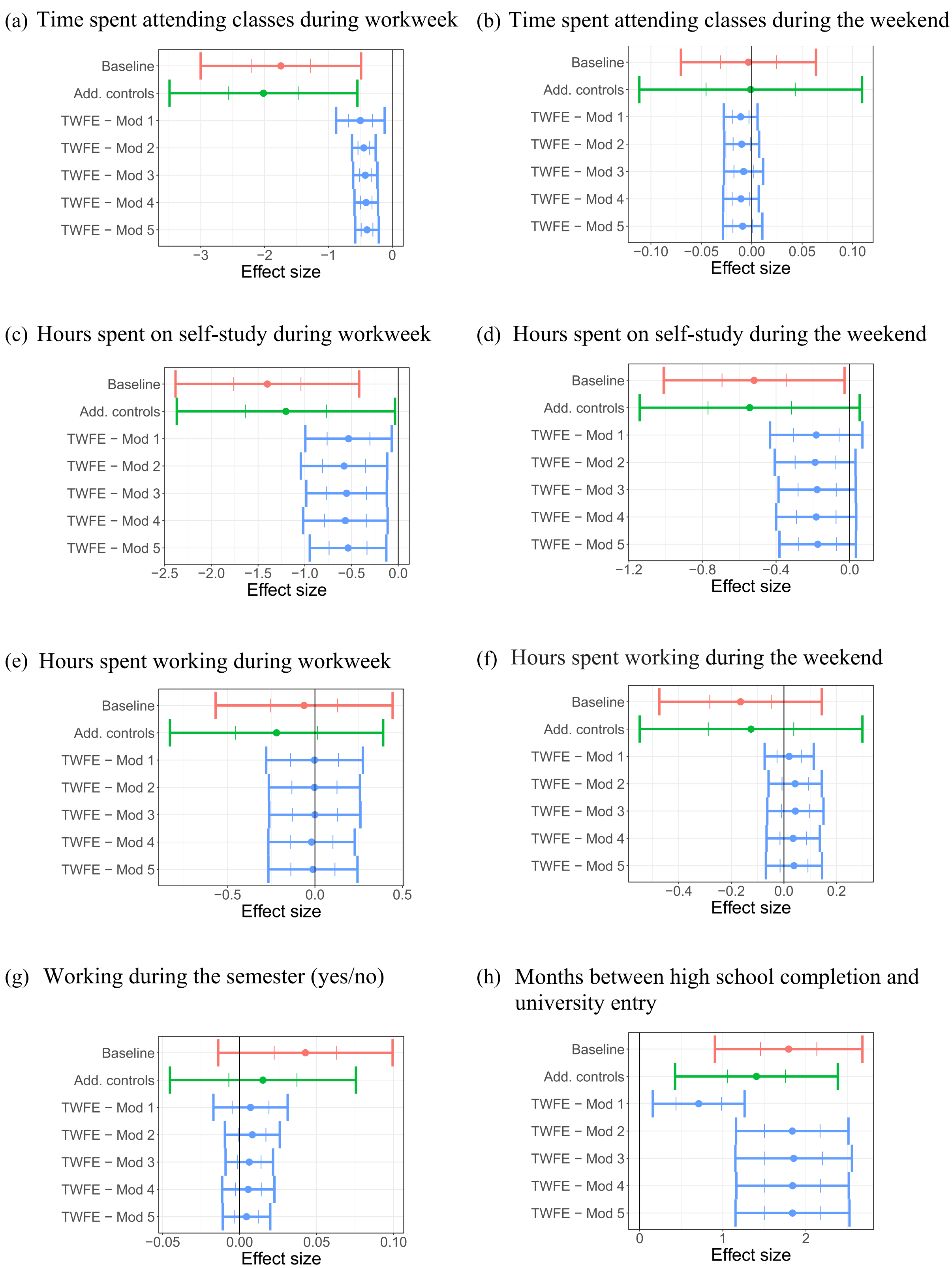

Fig. A1. Weighted Group-Time ATT and TWFE Estimates.

References

Abramitzky, R., Lavy, V., & Pérez, S. (2021). The long-term spillover effects of changes in the return to schooling. *Journal of Public Economics, 196*, Article 104369. https://doi.org/10.1016/j.jpubeco.2021.104369

Acemoglu, D., Autor, D. H., Hazell, J., & Restrepo, P. (2021). AI and jobs: Evidence from online vacancies. *NBER Working Paper Series AI*. https://doi.org/10.3386/w28257

Acemoglu, D., Lelarge, C., & Restrepo, P. (2020). Competing with robots: Firm-level evidence from France. *AEA Papers and Proceedings, 110*, 383–388. https://doi.org/10.1257/pandp.20201003

Acemoglu, D., & Restrepo, P. (2019). Automation and new tasks: How technology displaces and reinstates labor. *Journal of Economic Perspectives, 33*(2), 3–30. https://doi.org/10.1257/jep.33.2.3

Allensworth, E., Nomi, T., Montgomery, N., & Lee, V. E. (2009). College preparatory curriculum for all: Academic consequences of requiring algebra and English I for ninth graders in Chicago. *Educational Evaluation and Policy Analysis, 31*(4), 367–391. https://doi.org/10.3102/0162373709343471

Andersen, S. C., Humlum, M. K., & Nandrup, A. B. (2016). Increasing instruction time in school does increase learning. *Proceedings of the National Academy of Sciences of the United States of America, 113*(27), 7481–7484. https://doi.org/10.1073/pnas.1516686113

Andrietti, V. (2014). Does lecture attendance affect academic performance? Panel data evidence for introductory macroeconomics. *International Review of Economics Education, 15*, 1–16. https://doi.org/10.1016/j.iree.2013.10.010

Andrietti, V., & Su, X. (2019). The impact of schooling intensity on student learning: Evidence from a quasi-experiment. *Education Finance and Policy, 15*(4), 679–701. https://doi.org/10.1162/edfp_a_00263

Andrietti, V., & Velasco, C. (2015). Lecture attendance, study time, and academic performance: A panel data study. *Journal of Economic Education, 46*(3), 239–259. https://doi.org/10.1080/00220485.2015.1040182

Angrist, J. D., & Krueger, A. B. (1991). Does compulsory school attendance affect schooling and earnings? *Quarterly Journal of Economics, 106*(4), 979–1014. https://doi.org/10.2307/2937954

Angrist, J. D., & Pischke, J. S. (2009). Mostly harmless econometrics: An empiricist's companion. *An empiricist's companion*. Princeton university press. https://doi.org/10.1515/9781400829828

Apolinarski, B., Becker, K., Borchert, L., Bornkessel, P., Brandt, T., Fabian, G., Heißenberg, S., Isserstedt, W., Kandulla, M., Leszczensky, M., Link, J., Middendorff, E., Netz, N., Poskowsky, J., Schnitzer, K., Weber, S., & Wolter, A. (2021). 17th –21st Social Survey 2003 –2016. Data collection 2003, 2006, 2009, 2012, 2016. version: 1.0.0. data package access way: download-suF (E. Middendorff & M. Wallis (eds.)). https://doi.org/10.21249/DZHW:ssypool:1.0.1.

Arulampalam, W., Naylor, R. A., & Smith, J. (2012). Am I missing something? The effects of absence from class on student performance. *Economics of Education Review, 31*(4), 363–375. https://doi.org/10.1016/j.econedurev.2011.12.002

Athey, S., & Imbens, G. W. (2021). Design-based analysis in Difference-In-Differences settings with staggered adoption. *Journal of Econometrics, 226*(1), 62–79. https://doi.org/10.1016/j.jeconom.2020.10.012

Becker, R., & Schulze, A. (2013). In R. Becker, & A. Schulze (Eds.), *Bildungskontexte. Strukturelle Voraussetzungen und Ursachen ungleicher Bildungschancen*. Springer. https://doi.org/10.1007/978-3-531-18985-7

Bertrand, M., Duflo, E., & Mullainathan, S. (2004). How much should we trust differences-in-differences estimates? *Quarterly Journal of Economics, 119*(1), 249–275. https://doi.org/10.1162/003355304772839588

Bernardo, A., Esteban, M., Fernández, E., Cervero, A., Tuero, E., & Solano, P. (2016). Comparison of personal, social and academic variables related to university drop-out and persistence. *Frontiers in Psychology, 7*, 1–9. https://doi.org/10.3389/fpsyg.2016.01610

Biewen, M., & Schwerter, J. (2022). Does more maths and natural sciences in high school increase the share of female STEM workers? Evidence from a curriculum reform. *Applied Economics, 54*(16), 1889–1911. https://doi.org/10.1080/00036846.2021.1983139

Biewen, M., & Tapalaga, M. (2017). Life-cycle educational choices in a system with early tracking and "second chance" options. *Economics of Education Review, 56*, 80–94. https://doi.org/10.1016/j.econedurev.2016.11.008

Blanden, J. (2020). Education and inequality. In S. Bradley, & C. Green (Eds.), *The economics of education: A comprehensive overview*. Elsevier Ltd. https://doi.org/10.1016/B978-0-12-815391-8.00010-0

Bloom, B. S. (1968). Learning for mastery. *Evaluation Comment, 1*(2), 1854. https://doi.org/10.1007/978-1-4419-1428-6_4645

Bonesrønning, H., & Opstad, L. (2015). Can student effort be manipulated? Does it matter? *Applied Economics, 47*(15), 1511–1524. https://doi.org/10.1080/00036846.2014.997923

Borusyak, K., Jaravel, X., & Spiess, J. (2022). *Revisiting event study designs: Robust and efficient estimation* (pp. 1–54). http://arxiv.org/abs/2108.12419

Bourdieu, P. (1973). Cultural reproduction and social reproduction. In R. Brown (Ed.), *Knowledge, education, and cultural change: Papers in the sociology of education* (pp. 71–112). Tavistock Publication.

Bratti, M., & Staffolani, S. (2013). Student time allocation and educational production functions. *Annals of Economics and Statistics, 111/112*, 103–140. https://doi.org/10.2307/23646328

Büchele, S. (2020). Should we trust math preparatory courses? An empirical analysis on the impact of students' participation and attendance on short- and medium-term effects. *Economic Analysis and Policy, 66*, 154–167. https://doi.org/10.1016/j.eap.2020.04.002

Büttner, B., & 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. https://doi.org/10.1111/geer.12038

Callaway, B., & Sant'Anna, P. H. C. (2021a). *DID: Difference in differences. R package version 2.1.2*. https://bcallaway11.github.io/did/

Callaway, B., & Sant'Anna, P. H. C. (2021b). Difference-in-Differences with multiple time periods. *Journal of Econometrics, 225*(2), 200–230. https://doi.org/10.1016/j.jeconom.2020.12.001

Canning, E. A., LaCosse, J., Kroeper, K. M., & Murphy, M. C. (2020). Feeling like an imposter: The effect of perceived classroom competition on the daily psychological experiences of first-generation college students. *Social Psychological and Personality Science, 11*(5), 647–657. https://doi.org/10.1177/1948550619882032

Card, D. (1999). The causal effect of education on earnings. In D. C. Orley, & C. Ashenfelter (Eds.), *Handbook of labor economics* (3rd ed, pp. 1801–1863). Elsevier Science. https://doi.org/10.1016/S1573-4463(99)03011-4.

Carroll, J. B. (1988). The Carroll Model: A 25-Year Retrospective and Prospective View. *Educational Research, 18*(1), 26–31. https://doi.org/10.3102/0013189X018001026

Cortes, K. E., Goodman, J. S., & Nomi, T. (2015). Intensive math instruction and educational attainment: Long-run impacts of double-dose algebra. *Journal of Human Resources, 50*(1), 108–158. https://doi.org/10.3368/jhr.50.1.108

Dahmann, S. C. (2017). How does education improve cognitive skills? Instructional time versus timing of instruction. *Labour Economics, 47*, 35–47. https://doi.org/10.1016/j.labeco.2017.04.008

Dahmann, S. C., & Anger, S. (2014). The impact of education on personality: Evidence from a German high school reform. *SSRN Electronic Journal, 8139*. https://doi.org/10.2139/ssrn.2432423

DiPrete, T. A., & Eirich, G. M. (2006). Cumulative advantage as a mechanism for inequality: A review of theoretical and empirical developments. *Annual Review of Sociology, 32*, 271–297. https://doi.org/10.1146/annurev.soc.32.061604.123127

Dörsam, M., & Lauber, V. (2019). *The Effect of a compressed high school curriculum on university performance. Working paper series of the department of economics*. University of Konstanz, 2019–03 https://ideas.repec.org/p/knz/dpteco/1903.html

Duckworth, A. L., & Seligman, M. E. P. (2006). Self-discipline gives girls the edge: Gender in self-discipline, grades, and achievement test scores. *Journal of Educational Psychology, 98*(1), 198–208. https://doi.org/10.1037/0022-0663.98.1.198

Duckworth, A. L., Taxer, J. L., Eskreis-Winkler, L., Galla, B. M., & Gross, J. J. (2019). Self-Control and academic achievement. *Annual Review Psychology, 70*, 373–399. https://doi.org/10.1146/annurev-psych-010418-103230

Elder, G., Johnson, M., & Crosnoe, R (2003). The emergence and development of life course theory. In J. T. Mortimer, & M. J. Shanahan (Eds.), *Handbook of the life course* (pp. 3–19). Springer. https://doi.org/10.1007/978-0-306-48247-2_1

Ersoy, F. (2021). Returns to effort: Experimental evidence from an online language platform. *Experimental Economics, 24*(3), 1047–1073. https://doi.org/10.1007/s10683-020-09689-1

Fischer, A., & Roodman, D. (2021). *fwildclusterboot: Fast wild cluster bootstrap inference for linear regression models. R package (Version 0.13.0), 3*. https://cran.r-project.org/package=fwildclusterboot

Friedman-Sokuler, N., & Justman, M. (2016). Gender streaming and prior achievement in high school science and mathematics. *Economics of Education Review, 53*, 230–253. https://doi.org/10.1016/J.ECONEDUREV.2016.04.004

Gándara, D., & Li, A. (2020). Promise for whom? "Free-college" programs and enrollments by race and gender classifications at public, 2-year colleges. *Educational Evaluation and Policy Analysis, 42*(4), 603–627. https://doi.org/10.3102/0162373720962472

Gertler, P., Heckman, J., Pinto, R., Zanolini, A., Vermeersch, C., Walker, S., Chang, S. M., & Grantham-McGregor, S. (2014). Labor market returns to an early childhood stimulation intervention in Jamaica. *Science, 344*(6187), 998–1002. https://doi.org/10.1126/science.125117

Grave, B. S. (2011). The effect of student time allocation on academic achievement. *Education Economics, 19*(3), 291–310. https://doi.org/10.1080/09645292.2011.585794

Hackenberger, A., Rümmele, M., Schwerter, J., & Sturm, M. (2021). Elections and unemployment benefits for families: Did the Family Benefit Dispute affect election outcomes in Germany? *European Journal of Political Economy, 66*, Article 101955. https://doi.org/10.1016/j.ejpoleco.2020.101955

Hauschildt, K., Gwosć, C., Netz, N., & Mishra, S. (2015). Social and economic conditions of copyholders. *EUROSTUDENT V 2012–2015. wbv*. https://doi.org/10.3278/6001920bw

Hauschildt, K., Vögtle, E., & Gwosć, C. (2018). Social and economic conditions of student life in Europe. *EUROSTUDENT VI 2016–2018. wbv*. https://doi.org/10.3278/6001920cw

Heckman, J., Pinto, R., & Savelyev, P. (2013). Understanding the mechanisms through which an influential early childhood program boosted adult outcomes. *American Economic Review, 103*(6), 2052–2086. https://doi.org/10.1257/aer.103.6.2052

Henderson, D. J., Polachek, S. W., & Wang, L. (2011). Heterogeneity in schooling rates of return. *Economics of Education Review, 30*(6), 1202–1214. https://doi.org/10.1016/j.econedurev.2011.05.002

Henry, G. T., Pham, L. D., Kho, A., & Zimmer, R. (2020). Peeking into the black box of school turnaround: A formal test of mediators and suppressors. *Educational Evaluation and Policy Analysis, 42*(2), 232–256. https://doi.org/10.3102/0162373720908600

Hofmann, S., & 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. https://doi.org/10.1016/j.econedurev.2018.10.001

Homuth, C. (2017). Die G8-Reform in Deutschland [The G8-Reform in Germany]. *Otto-Friedrich-Universität Bamberg*. https://doi.org/10.1007/978-3-658-15378-6_2. Ph.D. thesis.

Hübner, N., Wagner, W., Kramer, J., Nagengast, B., & 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. https://doi.org/10.1007/s11618-017-0737-3

Hübner, N., Wagner, W., Meyer, J., & Watt, H. M. G. (2022). To those who have, more will be given? Effects of an instructional time reform on gender disparities in STEM subjects, stress, and health. *Frontiers in Psychology, 13*, 1–15. https://doi.org/10.3389/fpsyg.2022.816358

Huebener, M., Kuger, S., & Marcus, J. (2017). Increased instruction hours and the widening gap in student performance. *Labour Economics, 47*, 15–34. https://doi.org/10.1016/j.labeco.2017.04.007

Huebener, M., & Marcus, J. (2017). Compressing instruction time into fewer years of schooling and the impact on student performance. *Economics of Education Review, 58*, 1–14. https://doi.org/10.1016/j.econedurev.2017.03.003

Jacob, B., Dynarski, S., Frank, K., & Schneider, B. (2017). Are expectations alone enough? Estimating the effect of a mandatory college-prep curriculum in Michigan. *Educational Evaluation and Policy Analysis, 39*(2), 333–360. https://doi.org/10.3102/0162373716685823

Kemper, F.-J. (2008). Residential mobility in East and West Germany: Mobility rates, mobility reasons, reurbanization. *Zeitschrift für Bevölkerungswissenschaft, 33*(3–4), 293–314. https://doi.org/10.1007/s12523-009-0018-1

Kittelsen Røberg, K. I., & Helland, H. (2017). Do grades in higher education matter for labour market rewards? A multilevel analysis of all Norwegian graduates in the period 1990–2006. *Journal of Education and Work, 30*(4), 383–402. https://doi.org/10.1080/13639080.2016.1187265

Knol, M. J., Pestman, W. R., & Grobbee, D. E. (2011). The (mis)use of overlap of confidence intervals to assess effect modification. *European Journal of Epidemiology, 26*(4), 253–254. https://doi.org/10.1007/s10654-011-9563-8

Krashinsky, H. (2014). How would one extra year of high school affect academic performance in university? Evidence from an educational policy change. *Canadian Journal of Economics, 47*(1), 70–97. https://doi.org/10.1111/caje.12066

Kühn, S. M., Van Ackeren, I., Bellenberg, G., Reintjes, C., & Im Brahm, G. (2013). Wie viele Schuljahre bis zum Abitur? Eine multiperspektivische Standortbestimmung im Kontext der aktuellen Schulzeitdebatte. *Zeitschrift für Erziehungswissenschaft, 16*(1), 115–136. https://doi.org/10.1007/s11618-013-0339-7

Kulu, H., Mikolai, J., Thomas, M. J., Vidal, S., Schnor, C., Willaert, D., Visser, F. H. L., & Mulder, C. H. (2021). Separation and elevated residential mobility: A cross-country comparison. *European Journal of Population, 37*(1), 121–150. https://doi.org/10.1007/s10680-020-09561-1

Lavy, V. (2015). Do differences in schools' instruction time explain international achievement gaps? Evidence from developed and developing countries. *Economic Journal, 125*(588), 397–424. https://doi.org/10.1111/ecoj.12233

Lavy, V. (2021). The long-term consequences of free school choice. *Journal of the European Economic Association, 19*(3), 1734–1781. https://doi.org/10.1093/jeea/jvab001

Lehn, B. V. (2010). *Generation G8: wie die Turbo-Schule Schüler und Familien ruiniert [Generation G8: how the turbo school is ruining pupils and families]*. Beltz.

Lochner, L., Hanushek, E. A., Machin, S. J., & Woessmann, L. (2011). Nonproduction benefits of education: Crime, health, and good citizenship. In *Handbook of the economics of education* (pp. 183–282). Elsevier. https://doi.org/10.1016/B978-0-444-53444-6.00002-X

Lubinski, D., & Benbow, C. P. (2006). Study of mathematically precocious youth after 35 years: Uncovering antecedents for the development of math-science expertise. *Perspectives on Psychological Science, 1*(4), 316–345. https://doi.org/10.1111/j.1745-6916.2006.00019.x

MacKinnon, J. G., & Webb, M. D. (2017). Pitfalls when estimating treatment effects using clustered data. *The Political Methodologist, 24*(2), 20–31. https://www.econstor.eu/handle/10419/188899.

MacKinnon, J. G., & Webb, M. D. (2018). The wild bootstrap for few (treated) clusters. *Econometrics Journal, 21*(2), 114–135. https://doi.org/10.1111/ectj.12107

Marcus, J., Reif, S., Wuppermann, A., & Rouche, A. (2020). Increased instruction time and stress-related health problems among school children. *Journal of Health Economics, 70*, Article 102256. https://doi.org/10.1016/j.jhealeco.2019.102256

Marcus, J., & 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. https://doi.org/10.3368/jhr.54.2.1016.8324R

Marsh, H. W., & O'Mara, A. J. (2010). Long-term total negative effects of school-average ability on diverse educational outcomes: Direct and indirect effects of the big-fish-little-pond effect. *Zeitschrift für Pädagogische Psychologie, 24*(1), 51–72. https://doi.org/10.1024/1010-0652/a000004

Mayer, K. U. (2004). Whose lives? How history, societies, and institutions define and shape life courses. *Research in Human Development, 1*(3), 161–187. https://doi.org/10.1207/s15427617rhd0103_3

Metcalfe, R., Burgess, S., & Proud, S. (2019). Students' effort and educational achievement: Using the timing of the World Cup to vary the value of leisure. *Journal of Public Economics, 172*, 111–126. https://doi.org/10.1016/j.jpubeco.2018.12.006

Meyer, E., & Van Klaveren, C. (2013). The effectiveness of extended day programs: Evidence from a randomized field experiment in the Netherlands. *Economics of Education Review, 36*, 1–11. https://doi.org/10.1016/j.econedurev.2013.04.002

Meyer, T., & 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. https://doi.org/10.1086/684017

Meyer, T., & Thomsen, S. L. (2018). The role of high-school duration for university students' motivation, abilities and achievements. *Education Economics, 26*(1), 24–45. https://doi.org/10.1080/09645292.2017.1351525

Meyer, T., Thomsen, S. L., & Schneider, H. (2019). New evidence on the effects of the shortened school duration in the German States: An evaluation of post-secondary education decisions. *German Economic Review, 20*(4), 201–253. https://doi.org/10.1111/geer.12162

Milde-Busch, A., Blaschek, A., Borggräfe, I., von Kries, R., Straube, A., & Heinen, F. (2010). Is there an association between the reduced school years in grammar schools and headache and other health complaints in adolescent students? *Klinische Pädiatrie, 222*(4), 255–260. https://doi.org/10.1055/s-0030-1252012

Morin, L. P. (2013). Estimating the benefit of high school for university-bound students: Evidence of subject-specific human capital accumulation. *Canadian Journal of Economics, 46*(2), 441–468. https://doi.org/10.1111/caje.12019

Nomi, T., & Raudenbush, S. W. (2016). Making a success of "algebra for all": The impact of extended instructional time and classroom peer skill in Chicago. *Educational Evaluation and Policy Analysis, 38*(2), 431–451. https://doi.org/10.3102/0162373716643756

OECD. (2019). *OECD employment outlook 2019*. https://doi.org/10.1787/9ee00155-en

Oreopoulos, P. (2007). Do dropouts drop out too soon? Wealth, health and happiness from compulsory schooling. *Journal of Public Economics, 91*(11–12), 2213–2229. https://doi.org/10.1016/j.jpubeco.2007.02.002

Oreopoulos, P., & Salvanes, K. G. (2011). Priceless: The nonpecuniary benefits of schooling. *Journal of Economic Perspective, 25*(1), 159–184. https://doi.org/10.1257/jep.25.1.159

Patall, E. A., Cooper, H., & Allen, A. B. (2010). Extending the school day or school year: A systematic review of research (1985-2009). *Review of Educational Research, 80*(3), 401–436. https://doi.org/10.3102/0034654310377086

Pellizzari, M., & Billari, F. C. (2012). The younger, the better? Age-related differences in academic performance at university. *Journal of Population Economics, 25*(2), 697–739. https://doi.org/10.1007/s00148-011-0379-3

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* (523), 1216–1242. https://doi.org/10.1111/j.1468-0297.2007.02080.x

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, 238*(5), 441–476. https://doi.org/10.1515/jbnst-2018-0004

Ravitch, D. (2011). The death and life of the great American school system. *Early learning. Basic books*. https://doi.org/10.4324/9781315429977-13

Roodman, D., MacKinnon, J. G., Nielsen, M.Ø., & Webb, M. D. (2019). Fast and wild: Bootstrap inference in Stata using boottest. *Stata Journal, 19*(1), 4–60. https://doi.org/10.1177/1536867X19830877

Roth, J., Sant'Anna, P. H. C., Bilinski, A., & Poe, J. (2023). What's trending in difference-in-differences? A synthesis of the recent econometrics literature. *Journal of Econometrics, 235*(2), 2218–2244. https://doi.org/10.1016/j.jeconom.2023.03.008

Roth, T. (2019). What impact does the reduction in the overall length of the Gymnasium (G8) have on social inequalities in upper secondary school attendance? *Zeitschrift für Erziehungswissenschaft, 22*(5), 1247–1265. https://doi.org/10.1007/s11618-018-0847-6

Sánchez, A. C., & Andrews, D. (2011). Residential mobility and public policy. *OECD Journal: Economic Studies, 2011*, 185–206. https://doi.org/10.1787/eco_studies-2011-5kg0vswqt240

Scheerens, J. (1990). School effectiveness research and the development of process indicators of school functioning. *School Effectiveness and School Improvement, 1*(1), 61–80. https://doi.org/10.1080/0924345900010106

Schwerter, J., Dimpfl, T., Bleher, J., & Murayama, K. (2022a). Benefits of additional online practice opportunities in higher education. *Internet and Higher Education, 53*, Article 100834. https://doi.org/10.1016/j.iheduc.2021.100834

Schwerter, J., Wortha, F., & Gerjets, P. (2022b). E-Learning with multiple-try-feedback: Can hints foster students' achievement during the semester? *Educational Technology Research and Development, 70*, 713–736. https://doi.org/10.1007/s11423-022-10105-z

Stinebrickner, R., & Stinebrickner, T. R. (2008). The causal effect of studying on academic performance. *The B.E. Journal of Economic Analysis and Policy, 8*(1). https://doi.org/10.2202/1935-1682.1868

Sun, L., & Abraham, S. (2021). Estimating dynamic treatment effects in event studies with heterogeneous treatment effects. *Journal of Econometrics, 225*(2), 175–199. https://doi.org/10.1016/j.jeconom.2020.09.006

Taylor, S. E. (2010). Mechanisms linking early life stress to adult health outcomes. *Proceedings of the National Academy of Sciences of the United States of America, 107* (19), 8507–8512. https://doi.org/10.1073/pnas.1003890107

Thiel, H., Thomsen, S. L., & 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. https://doi.org/10.1111/rssa.12079

Tymms, P., Merrell, C., & Bailey, K. (2018). The long-term impact of effective teaching. *School Effectiveness and School Improvement, 29*(2), 242–261. https://doi.org/10.1080/09243453.2017.1404478

Verbree, A. R., Hornstra, L., Maas, L., & Wijngaards-de Meij, L. (2023). Conscientiousness as a predictor of the gender gap in academic achievement. *Research in Higher Education, 64*(3), 451–472. https://doi.org/10.1007/s11162-022-09716-5

Winkelmann, L., & Winkelmann, R. (1998). Why are the unemployed so unhappy? Evidence from panel data. *Economica, 65*, 1–15. https://doi.org/10.1111/1468-0335.00111

Zimmer, R., Henry, G. T., & Kho, A. (2017). The effects of school turnaround in tennessee's achievement school district and innovation zones. *Educational Evaluation and Policy Analysis, 39*(4), 670–696. https://doi.org/10.3102/0162373717705729