Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

How Shorter Schooling Shapes Family Formation: Impacts from Germany's G8 Reform

Leonhard Friedel

leonhard.friedel@whu.edu

WHU – Otto Beisheim School of Management

Talha Umar

WHU – Otto Beisheim School of Management

Research Article

Keywords: Education System, Fertility, Germany, G8, Educational Policy

Posted Date: November 23rd, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11612146/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

How Shorter Schooling Shapes Family Formation: Impacts from Germany's G8 Reform

Leonhard Friedel[1,2] **and Talha Umar**[1]

Abstract

We analyze the causal impact of Germany's G8 reform, which shortened Gymnasium from nine to eight years, on fertility. Using longitudinal data from the German SOEP, we link individuals' graduation cohorts and states of schooling to identify exposure. Exploiting the staggered implementation across states, we apply a stacked generalized difference-in-differences design following Cengiz et al. (2019), controlling for socioeconomic characteristics. Our results show that exposure to the reform significantly reduces the total number of children born, with stronger effects in East Germany and among university graduates nationwide. The findings suggest that accelerated schooling compresses early life-course transitions, raises opportunity costs of family formation, and ultimately lowers fertility, offering complementary insights for demographic and education policy.

Keywords: Education System; Fertility; Germany; G8; Educational Policy

Leonhard Friedel[1,2]: Leonhard.Friedel@whu.edu | ORCID: 0009-0008-9716-6387
Talha Umar[1]: Talha.Umar@whu.edu | ORCID: 0009-0000-4604-8403

Declarations of interest: none

[1] WHU – Otto Beisheim School of Management, Vallendar (Germany)

[2] Corresponding author

The authors received no funding for this research.

1. Introduction

Since 1971, Germany has experienced persistently low fertility rates, consistently falling below the replacement level. In 2019, the fertility rate stood at a notably low 1.51 children per woman. This study investigates whether a one-year reduction in the duration of schooling due to a decentralized education reform influenced the fertility rate of women up to age 35. Specifically, we examine the impact of shortening the time required to obtain the Abitur, Germany's university entrance qualification, on family formation decisions. Despite the existence of a mostly uniform national framework, education in Germany is governed by state authorities, resulting in variations in school systems across federal states.

Historically, all West German states required nine years of education in Gymnasium to earn the Abitur. Following the unification of Germany and subsequent educational reforms, as part of the more recent G8 reform, many states changed the academic secondary school track from nine to eight years while maintaining the curriculum. This study leverages these differences as a natural experiment to explore the relationship between reduced schooling duration and the fertility rate.

Empirically, we exploit the staggered rollout of G8 across German federal states and cohorts. Since education policy in Germany is decentralized, the first Abitur cohorts exposed to G8 differ by state and year, generating quasi-experimental variation in treatment exposure. We combine two data sources: (i) individual-level, longitudinal information from the German Socio-Economic Panel (SOEP), which provides fertility histories and rich socio-economic covariates; and (ii) administrative information from state education ministries on the timing of G8 implementation by state. Our analysis focuses on women who obtained the Abitur and are observed in SOEP; we harmonize individuals by state of schooling and Abitur cohort to assign treatment status, and we study fertility outcomes up to age 35 to concentrate on the main reproductive window for highly educated women and to limit lifecycle-incompleteness concerns.

The paper speaks to a large literature on education and fertility (and education and demographic outcomes more broadly) that has largely focused on reforms changing *attainment*, such as compulsory schooling laws and expansions of educational access. A robust finding in many contexts is that additional education reduces very early and teenage childbearing, while effects on completed fertility and childlessness are smaller and context-dependent. At the same time, evidence from Germany and other European settings highlights the central role of opportunity costs, career timing, and institutional constraints for highly educated women's fertility trajectories. By studying a reform that primarily changes when education is completed rather than how much education is obtained, we provide evidence that complements the interpretation of this prior work: the timing of educational progression itself can matter for family formation, even while holding the curriculum fixed.

We make three contributions related to this literature. First, we extend the literature on education and fertility by exploiting a reform that altered the duration and timing of academically oriented secondary schooling while holding formal attainment constant. This stands in contrast to compulsory schooling reforms, which typically raise minimum education levels and confound timing effects with changes in qualifications. Second, we provide novel evidence on how such a reform affects fertility in a setting characterized by high opportunity

costs of childbearing and persistent institutional heterogeneity, namely, East and West Germany. Third, the staggered implementation of the G8 reform allows us to employ a stacked difference-in-differences framework that not only identifies the causal effect of shortening the Gymnasium track but also, through a time-reversed specification, enables us to test the counterfactual effect of lengthening it as an additional robustness check.

Our main results show that exposure to G8 reduced total fertility by the mid-thirties. The average effect is modest but economically meaningful, with particularly pronounced reductions in East Germany. The complementary G9 counterfactual indicates that extending the Gymnasium track would have increased fertility, with larger effects in East Germany and among university graduates. These patterns suggest that the compression of upper secondary schooling affects the timing and feasibility of early family formation, and that the magnitude of these effects is shaped both by regional institutional environments and by the opportunity costs faced by highly educated women.

The remainder of the paper is structured as follows. Section 2 reviews the related literature. Section 3 presents the data. Section 4 outlines the methodology and identification strategy. Section 5 reports the empirical results. Section 6 discusses their interpretation in light of potential mechanisms, existing theories, and the German institutional context. Section 7 concludes with implications for education and family policy.

2. Literature Review

The German school system consists of elementary education followed by three main tracks. After completing four years of primary education ('Grundschule'), students have the option to enroll in one of three tracks: 'Hauptschule', 'Realschule', and 'Gymnasium'. The Hauptschule provides basic general education, usually up to grade 9, and mainly prepares students for entry into the dual system of vocational education and training. The Realschule offers a broader academic curriculum, typically up to grade 10, and leads to the Mittlere Reife, which qualifies students for many white-collar vocational programs and some further schooling but does not, on its own, confer university access. The Gymnasium represents the academic track that prepares students for the Abitur, the general university entrance qualification. In addition, many German states operate comprehensive schools (Gesamtschulen) that combine these tracks within a single institution, although the tripartite structure remains the defining feature of German secondary education.

The Gymnasium track has traditionally required nine years to complete (hence called G9), resulting in 13 years of total schooling. Two exceptions of note here are Saxony and Thuringia, which have maintained the 8-year Gymnasium (G8) system since 1949. As part of the G8 reform, the remaining federal states of Germany moved to the 8-year track between 2001 and 2007, resulting in the first students graduating under the new system between 2007 and 2016. The intent behind the reform was to bring Germany's school leaving age in line with other countries, enabling graduates to enter the labor market earlier and increasing their international competitiveness. Notably, as part of the reform, the study requirements of the Abitur were not changed and instead, the G9 curriculum was compressed into 8 years (Huebener & Marcus, 2017).

The impact of the reform has been studied by Huebener and Marcus (2017), who note that while the reform resulted in an increase of approximately 3.6% in students repeating grades, the overall graduation rate remained unchanged and graduates of the G8 program were on average 10 months younger than their G9 counterparts. Subsequent contributions find that G8 students report higher levels of stress and lower subjective well-being than comparable G9 students (Hofmann & Mühlenweg, 2018; Hübner et al., 2017) with minor to no effect on cognitive outcomes (Hübner et al., 2017; Meyer & Thomsen, 2018; Quis, 2018). The negative effects also carry over to university enrollment, with a 10% reduction among women and a negligible effect among men in the first year (Meyer et al., 2019). Marcus and Zambre (2019) extend this by finding a 7.3% reduction in overall enrollment in the first year with a 9.6% higher probability of delaying enrollment.

Overall, this literature suggests that the G8 reform reduced the age at graduation while leaving graduation rates and average academic achievement largely intact, at the cost of increased grade repetition, higher perceived stress and some deterioration in subjective health and leisure outcomes. Furthermore, while the effect does not translate fully, it also reduced the age of university graduates.

The broader literature covering the impact of increased education on first childbearing, total lifetime fertility, and childlessness has offered mixed insights on the topic. Focusing on Europe and the US, while most studies find a reduction in the incidence of teenage births (Black et al., 2008; Monstad et al., 2008; Silles, 2011), there is little to no evidence on the reduction in average children and childlessness (Monstad et al., 2008). Fort et al. (2016) find a negative relationship between education and fertility in the UK but this relationship does not hold true for Continental Europe. DeCicca & Krashinsky (2023) study education reforms in Canada and find that increased education "compresses" the fertility distribution: it leads to decreased childlessness while also decreasing total fertility.

West and East Germany retained structurally distinct family formation regimes well beyond reunification, reflecting persistent institutional and normative differences that matter for the evaluation of education policies. Prior to 1990, the GDR was characterized by early and nearly universal entry into parenthood, low levels of childlessness, and a stable two-child norm, supported by extensive public childcare provision and high female full-time employment (Kreyenfeld, 2003, 2004). Following reunification, East German women shifted the timing of first births toward West German ages but continued to exhibit higher first-birth intensities and substantially lower second-birth transitions (O. A. Becker et al., 2010; Goldstein & Kreyenfeld, 2011). To our knowledge, no study has examined fertility responses to education reforms in the former GDR. For West Germany, evidence from the extension of compulsory schooling from eight to nine years between 1946 and 1969 shows that more education reduced fertility: Cygan-Rehm & Maeder (2013) document delayed first births without subsequent catch-up, resulting in lower completed fertility and a higher likelihood of childlessness. This is attributed to the high opportunity costs of raising offspring in Germany. For the effects of compressing time while maintaining curriculum, several studies exploit the Short School Year reform of 1966-67 and find earlier family formation with no effect on lifetime fertility rates (Koebe & Marcus, 2022) and no adverse effects on human capital acquisition (Pischke, 2007).

Vasireddy et al. (2023) review several mechanisms linking education to fertility. Building on Becker's economic theory, children can be understood as normal goods, so desired fertility may rise with income, but is constrained by the direct and indirect costs of childrearing (G. S.

Becker, 1960; G. S. Becker and Lewis, 1973; Willis, 1973). Direct costs include spending on housing, food, childcare and education, while indirect costs include forgone earnings and career opportunities, which can be substantial, particularly for mothers (Joshi, 1990). Because education raises wages and career prospects, it increases the opportunity cost of childbearing and may encourage fewer, higher-investment children, especially where support for combining work and family is limited. At the same time, education can increase the likelihood of having any children by raising permanent income, improving partnership prospects and enabling better fertility planning. Empirical evidence therefore suggests that education often delays and compresses childbearing, reducing teenage and very early motherhood, while having more moderate or context-dependent effects on completed fertility (Black et al., 2008; DeCicca and Krashinsky, 2023; Fort et al., 2016; Monstad et al., 2008; Vasireddy et al., 2023). Finally, because female fecundity declines with age, especially after the mid-thirties, longer education may also shift childbearing into ages where conception is more difficult and involuntary childlessness becomes more likely. The G8/G9 reform differs conceptually from the aforementioned reforms: it alters the timing and intensity of academically oriented secondary education without changing the minimum compulsory schooling requirement. As such, it provides a complementary quasi-experiment that mainly shifts the age profile of schooling and transitions into tertiary education and the labor market. Hence, holding all other life-course stages constant, we examine whether reducing schooling by one year effectively extends the period of viable fertility by an additional year. If earlier completion of education allows women to enter the phase of family formation sooner, this shift could relax the biological time constraints they face. Consequently, such an extension may lead to higher fertility rates, as women gain an additional year during which conception is more likely from a biological standpoint.

3. Data

Our analysis is based on a representative cross-section of the German population for each year, focusing on individuals who graduated from a Gymnasium (academic-track secondary school). The primary data source is the German Socio-Economic Panel (SOEP, 2024), a nationally representative longitudinal survey of private households and individuals in Germany.

To account for the fact that the SOEP does not perfectly trace individuals over time, we use the last available wave for each individual from 1988 onwards. We extract the relevant information for each individual and continue our analysis using individual-level data organized by birth cohorts.

To identify when the G8 reform, which shortened the duration of Gymnasium from nine to eight years, took effect across German states, we draw on administrative data from the respective state ministries of education. The reform was implemented at different times across the federal states, creating regional variation in its timing. Saxony-Anhalt was the first to implement the reform, with the first G8 graduates finishing in 2007, followed by Mecklenburg-Western Pomerania in 2008 and Saarland in 2009. Hamburg introduced G8 in 2010, while Bavaria and Lower Saxony followed in 2011. A larger group of states including Baden-Württemberg, Berlin, Brandenburg, and Bremen shifted to the shortened school track in 2012. Hesse and North Rhine-Westphalia adopted the reform in 2013, and Schleswig-Holstein and

Rhineland-Palatinate were the last to implement it, with their first G8 graduates in 2016 (see Table 1). Importantly, Saxony and Thuringia already had an eight-year Gymnasium system in place before the nationwide G8 reform efforts and thus were not affected by these changes.

[Insert Table 1 here]

Table 1 provides descriptive statistics for the main variables used in the empirical analysis. The summary statistics refer to the stacked dataset constructed for the evaluation of the G8 reform and include all female respondents who completed the Gymnasium track (Abitur) and are observed in their last available SOEP interview since 1988.

[Insert Table 2 here]

The outcome variable, *TFR*, is the observed total fertility rate, adjusted for the observation period of the dataset, since individuals are observed over a varying number of years. The mean of 0.23 reflects the relatively young age profile of the sample; the maximum value indicates the presence of a small number of high-parity cases. The key treatment indicator *G8_post_stack* captures whether an individual belongs to a cohort that completed secondary schooling under the shortened eight-year Gymnasium track. Approximately 14 percent of observations fall into this category, consistent with the staggered implementation of the reform across federal states. Related dummy variables indicate whether the observation belongs to a post-reform cohort in East Germany (*G8_east*) or whether the individual later obtained a university degree (*G8_and_Uni*).

The table also reports summary statistics for a series of socio-demographic and labor-market covariates used in the regression analysis. *Net_income* reflects monthly net labor income at the time of the last SOEP interview and *Publ_servant* indicates employment in the public sector. *Married* captures marital status. Measures of education and vocational training are included through variables such as *Apprenticeship*, recorded as the cumulative number of vocational certificates obtained.

The dataset further contains identifiers used in the stacked event-study design: *stack_state* denotes the normalized state identifier within each stack, *rel_year* measures the distance (in years) to the reform implementation with negative values for pre-treatment cohorts, and *rel_year_pos* shifts this variable to strictly positive values.

Together, these statistics illustrate (i) substantial variation in exposure to the G8 reform, (ii) meaningful sociodemographic heterogeneity across individuals, and (iii) the balanced cohort structure required for the stacked difference-in-differences design applied in Section 4.

4. Methodology

To estimate the causal effect of the G8 high school reform on fertility outcomes in Germany, we employ a stacked (dynamic) difference-in-differences (DiD) design following the design of Cengiz et al. (2019), which fulfills the same requirements for consistency under heterogeneous

treatment effects as newer estimators like Callaway and Sant'Anna (2021). By constructing clean cohort-specific sub-experiments, both approaches mathematically eliminate the 'forbidden comparisons' inherent in standard two-way fixed effects (TWFE) regressions. In Figure 1, we show a simplified version of our data set. We have multiple groups, which represent the different states over different time periods (years) with respective treatment in black (the implementation of the G8 regime), while the cohorts under the G9 regime are represented in gray. Some states always had a G8-regime (Group A), additionally we have early adopters (B), adopters (C) and late adopters (D). Finally, some states never implemented the G8 regime during the sample period, as represented by Group E.

[Insert Figure 1 here]

To capture the causal effect of the implementation of the G8 high school reform on fertility, we employ an established approach and an unconventional robustness check. Both are designed to assess the impact of reduced schooling duration on the number of children born.

First, we apply the standard stacked regression approach and restrict the sample to individuals who finished secondary school between 2000 and 2015 and obtained a university entrance qualification (Abitur). We exclude Saxony and Thuringia from the analysis because they always operated under the G8 system, similar to group A (always treated) in Figure 1. We then apply a stacking procedure that redefines time relative to each state's reform date. Rhineland-Palatinate and Schleswig-Holstein serve as control states, as they consistently kept a 9-year school duration over the entire sample period (2000–2015), comparable to group E in Figure 1. In Figure 1, assume for illustration that treatment begins before Period 3. In the stacked design, group C is shifted one period back in time and group D is shifted two periods back, while group E serves as a control group that is duplicated for each stack and moved backward accordingly. After constructing these stacked datasets, we can evaluate the G8 reform as a standard difference-in-differences (DiD) policy intervention using a two-way fixed effects (TWFE) specification.

In our robustness check, we exploit the staggered rollout of the reform together with the fact that Saxony and Thuringia always kept the 8-year secondary school system, just like group A in Figure 1. This setting allows us to use a time-inversed stacked difference-in-differences (DiD) design. We treat the G9 system as a hypothetical policy intervention and estimate its fertility effects within the treated states, while Saxony and Thuringia serve as stable control groups. The time-inversed setup lets us apply a standard two-way fixed effects (TWFE) model and identify effects under the usual common-trends assumption. However, this approach risks what Wang et al. (2024) call the "forbidden regression," in which pre-treated units are later used as non-treated units. This is problematic if treatment effects are heterogeneous over time, for example if effects accumulate the longer a reform has been in place, which can bias the estimates. In addition, spillover effects from already treated units to later treated units could distort the results.

Given our data, the cohort structure, and especially the short time horizon, we argue, based on biological rather than economic or statistical reasoning, that these concerns are unlikely to apply here. We compare different cohorts of individuals who are independent of previous cohorts in terms of treatment status. The panel dimension comes from observing multiple states

over time, where each individual in a state is either treated or not in a given period. For spillovers or other heterogeneous effects to change fertility patterns in a way that feeds back onto previously treated groups, children would first need to grow up and reach reproductive age. This would typically take more than 18-21 years, which is longer than the time span covered by our data.

To be specific, we design our stacked difference-in-differences according to the following specifications:

$$rel_year_{ist} = (year_of_graduation_{it} - reform_year_{s})$$

The variable *rel_year* serves as the new time variable, centered around the policy intervention. To account for the transition from G9 to G8, where rel_year = 0 marks the year of implementation, we drop this cohort to prevent attenuation bias, since in the first reform year both G8 and G9 graduates are observed simultaneously due to double graduation cohorts (Doppeljahrgänge), making treatment status non-identifiable for this group. The indices *i*, *s*, and *t* denote individuals, states, and graduation (calendar) years, respectively.

To obtain a strictly non-negative time variable, we construct *Rel_Time* by shifting rel_year_{ist} such that its minimum value equals 1. Specifically, we define:

$$Rel_Time_{ist} = rel_year_{ist} + |min(rel_year_{ist})| + 1$$

Next, we define a binary indicator *Post* to indicate if we are before or after the intervention. *Post* equals 1 if the value of *Rel_Time* exceeds $|min(rel_yearist)| + 1$ (i.e., if the observation occurs after the reference period), and 0 otherwise:

$$Post = 1 \text{ if } Rel_Time_{ist} > |min(rel_year_{ist})| + 1, \text{ and 0 otherwise}$$

This construction ensures that *Rel_Time* is strictly positive and that *Post* accurately captures the post-intervention period relative to the normalized timeline. The treatment indicator is defined at the state level and equals 1 for states that implemented the policy intervention, and 0 otherwise. Finally, each stack is built following the standard procedure following Cengiz et al. (2019).

We construct the dependent variable as the number of children ever born to a woman who has completed, or is expected to complete, the Abitur. To this end, we divide the number of children by the number of years between the year of (expected) Abitur completion and the survey year to obtain an average number of births per year.

Based on this preparation, we apply a simple TWFE difference-in-differences estimation. In the main approach, we restrict the dataset to individuals who graduated between 2000 and 2015 and exclude states that have historically implemented only the G8 system, namely Thuringia and Saxony. The G8 policy implementation is treated as the intervention. Accordingly, the treatment variable equals 1 for states that implemented the G8 reform and 0 otherwise. Given that the dataset ends in 2015, Rhineland-Palatinate and Schleswig-Holstein, which retained the G9 system throughout the observation period, serve as clean control states.

Specification I:

$$TFR_{ist} = \beta_1 \bullet Post_t \times Treatment_s + \beta_3 \bullet Controls_{ist} + \mu_s + \eta_t + \gamma_c + \varepsilon$$

Specification II:

$$TFR_{ist} = \beta_1 \bullet Post_t \times Treatment_s + \beta_2 \bullet Post_t \times Treatment_s \times East_{st} + \beta_3 \bullet Controls_{ist} + \mu_s + \eta_t + \gamma_c + \varepsilon$$

Specification III:

$$TFR_{ist} = \beta_1 \bullet Post_t \times Treatment_s + \beta_2 \bullet Post_t \times Treatment_s \times Graduate_{ist} + \beta_3 \bullet Post_t \times Treatment_s \times East_{st} + \beta_4 \bullet Controls_{ist} + \mu_s + \eta_t + \gamma_c + \varepsilon$$

The dependent variable, *TFR*, represents the observed total fertility rate, adjusted for variation in the number of years over which individuals are observed. The interaction term *Post* × *Treatment* identifies the average effect of the G8 reform in the post-implementation period for treated states. The vector of control variables includes relevant individual-level characteristics such as net income, marital status, educational level, and occupational position. To account for unobserved heterogeneity, we include state fixed effects (μ_s), stack-time fixed effects (η_t), and birth cohort fixed effects (γc). The error term ε absorbs any remaining unexplained variation. In specification II, we use the additional triple interaction term *Post* × *Treatment* × *East* to capture the treatment effect on the post-Soviet states of the former German Democratic Republic (GDR), which joined the Federal Republic of Germany (FRG) in 1990. Although reunification took place more than three decades ago, a clear division between East and West Germany persists to this day with regard to culture, economic activity, infrastructure, and social structures. In specification III, we add the additional interaction term *Post* × *Treatment* × *Graduate*, which allows for differential effects based on educational attainment, where Graduate is a binary indicator equal to one for individuals who have completed a university degree. The coefficient on this triple interaction term captures the additional impact of the reform for this subgroup. This specification enables a consistent and robust estimation of both the average and heterogeneous effects of the G8 reform on fertility behavior, while controlling for a wide range of confounding factors. We cluster standard errors at the state level in all specifications.

For the robustness check, we mainly follow the same strategy, except we invert the time axis, do not restrict the data set from 2000-2015, and include the states Thuringia and Saxony with their permanent G8 regimes as clean controls. The previous control group states are treated since 2016 and therefore in this approach part of the treatment group. To reflect the change from G8 to G9 (i.e., the policy reversal of the second approach), we multiply *rel_year* by -1, so that *rel_year = 1* corresponds to the first year after the reversion.

In the robustness check, we apply the following specifications.

Specification IV:

$$TFR_{ist} = \beta_1 \bullet Post_t \times Treatment_s + \beta_3 \bullet Controls_{ist} + \mu_s + \eta_t + \gamma_c + \varepsilon$$

The interaction term *Post* × *Treatment* captures the treatment effect. The variables used in these specifications mirror those defined in the first approach. In Specification V, we include an interaction term between the treatment effect (*Post* × *Treatment*) and an indicator for East German states. This allows us to assess whether the treatment effect differs between the East and West German states, capturing potential heterogeneity in response to the policy intervention.

Specification V:

$$TFR_{ist} = \beta_1 \bullet Post_t \times Treatment_s + \beta_2 \bullet Post_t \times Treatment_s \times East_{ist} + \beta_3 \bullet Controls_{ist} + \mu_s + \eta_t + \gamma_c + \varepsilon$$

In Specification VI, we additionally include an interaction term between the treatment effect (*Post* × *Treatment*) and an indicator for individuals who hold a university degree.

Specification VI:

$$TFR_{ist} = \beta_1 \bullet Post_t \times Treatment_s + \beta_2 \bullet Post_t \times Treatment_s \times Graduate_{ist} + \beta_3 \bullet Post_t \times Treatment_s \times East_{ist} + \beta_4 \bullet Controls_{ist} + \mu_s + \eta_t + \gamma_c + \varepsilon$$

The extended specification allows for the treatment effect to vary by educational attainment. Specifically, we interact the main treatment indicator (*Post* × *Treatment*) with a binary variable *Graduate* indicating whether an individual has completed a university degree. This enables us to capture potential heterogeneity in fertility responses to the reform between those with and without tertiary education. In this model, the coefficient on the main interaction term (*Post* × *Treatment*) reflects the treatment effect for individuals without a university degree, while the coefficient on the triple interaction term captures the additional effect for university graduates. By including this interaction, we account for the possibility that individuals with higher levels of education may respond differently to changes in the timing and structure of educational trajectories induced by the G8 or hypothetical G9 reform. All other covariates and fixed effects remain as specified in the baseline model, ensuring comparability of estimates across specifications.

Common Trend assumption

In both empirical approaches, we assess the validity of the common trends assumption, which is central to the causal interpretation of the difference-in-differences estimates. We proceed in several steps, starting with a simple comparison of treatment and control groups and then

examining their pre-reform dynamics in more detail. First, we test whether treatment and control groups differ systematically before the reform. Table A2 reports this common trend test for the main G8 specification and the G9 robustness check. The treatment/control coefficient captures the average pre-reform difference between the two groups, conditional on stack-state and stack-time fixed effects. Neither coefficient is statistically significant, providing no evidence of a systematic pre-reform difference.

We then conduct an event-study analysis to examine whether the groups followed different trends over time. Figure A1 presents the G8 results and Figure A2 the corresponding G9 results. The pre-treatment coefficients are not different from zero, show no systematic trend, and none is individually statistically significant. We additionally test whether the pre-treatment coefficients are jointly equal to zero. Conventional joint F-tests yield $F(6,11) = 2.88$, $p = 0.062$ for the main approach and $F(6,11) = 0.99$, $p = 0.478$ for the robustness check. The p-value of 0.062 for the main approach is close to conventional significance levels and could therefore raise concerns about a possible pre-trend. However, given the small number of clusters (12 states), conventional cluster-robust inference can be unreliable and may overstate statistical significance. We therefore additionally conduct wild cluster bootstrap tests using Webb weights and 9,999 replications, which are better suited to settings with a small number of clusters. The resulting p-values increase substantially to 0.555 for the main approach and 0.914 for the robustness check, suggesting that the main approach's near-significance under the conventional test likely reflects the small-cluster asymptotic approximation rather than evidence of a systematic pre-trend. Taken together, the results from Table A2 and the event-study analyses in Figures A1 and A2 support the validity of the parallel trends assumption underlying our difference-in-differences design.

5. Results

Table 3, Table 4, and Table 5 report results from the stacked difference-in-differences (DiD) approaches that examine the relationship between school duration reforms and the number of children. The main approach estimates the actual effect of the G8 reform, which shortened Gymnasium by one year. The robustness check uses a time-inversed DiD setup to approximate the hypothetical effect of adding one year of schooling (G9).

[Insert Table 3 here]

In Table 3, which focuses on the realized G8 reform, the treatment effects are consistently negative and statistically significant in Specification I across all three models, with magnitudes ranging from –0.275 to –0.133 children on average. This indicates that the introduction of G8 is associated with a reduction in the number of children. The baseline model (Specification I, Model 1) shows the largest negative effect of the shortened school duration; the size of the coefficient decreases once covariates are added.

Specification II introduces an interaction for East German states. The results (particularly in Model 4) show that the strongest decline in the number of children due to the reform occurs in East Germany. When controlling for marital status, net income, public sector employment, and detailed educational qualifications, the negative impact of the reform can largely be attributed to the East German states.

Across both specifications, several covariates follow expected patterns. Marriage exhibits a strong and highly significant positive association with the number of children. Certain vocational qualifications, in particular apprenticeships, healthcare-related vocational training and healthcare schooling completed before 1999, are also positively correlated with fertility. Other types of non-academic training, such as civil service training or technical college degrees, do not show a significant relationship. Furthermore, neither net income nor employment as a state officer has a meaningful effect on fertility.

Although university degrees are not included explicitly as covariates in this specification, the pattern observed in the job and qualification variables suggests that higher educational attainment is likely associated with a lower number of children. The positive fertility correlations appear primarily among vocationally trained individuals, while occupations and training pathways characteristic of higher educational trajectories do not show similar effects.

[Insert Table 4 here]

Table 4 reports the robustness test results from the time-inversed stacked DiD design; the estimates approximate the hypothetical effect of extending school duration by one year through a G9 treatment indicator. In Specification IV, the treatment effects are positive and statistically significant across the three models, with coefficients between 0.169 and 0.195 children. These results mirror the findings from Table 3 in reversed form: whereas the realized G8 reform reduced fertility, the hypothetical extension to G9 is associated with an increase in the number

of children. Similar to Table 3, the direction of the effect remains consistent when covariates are added.

Specification V (Table 4) introduces an interaction for East German states. In this specification, the interaction term is sizeable and becomes statistically significant once covariates are added, with coefficients of 0.321 and 0.319 (Models 5 and 6). This indicates that an extension of schooling duration would have had the strongest fertility-increasing effect in East Germany. Similar to the pattern observed in the G8 estimates, regional heterogeneity plays an important role: once East–West differences and covariates are taken into account, the fertility response to a longer school duration appears to be driven predominantly by the eastern states.

The covariates behave largely in line with the estimates from Table 3. Marriage again shows a strong and significant positive association with the number of children. Vocational qualifications also display similar patterns: apprenticeships are positively related to fertility, whereas other vocational and non-academic education categories do not yield significant effects. Net income and employment as a state officer continue to show no meaningful relationship with the number of children.

Taken together, the findings from Table 4 reinforce the interpretation from Table 3. Shortening school duration (G8) is associated with lower fertility, while lengthening school duration (G9) produces positive fertility effects, particularly in East Germany. These aligning patterns further support the idea that the observed effects are not artifacts of the modeling strategy but reflect meaningful variation in how school duration interacts with regional labor market conditions, educational trajectories, and family formation behavior. As university education does not appear as a separate control in these specifications, the next step is to include an interaction term for university graduates to examine whether these effects differ across educational groups.

[Insert Table 5 here]

Table 5 extends the analysis by interacting the treatment indicators with a dummy for university graduates to capture heterogeneity in fertility responses across educational groups. In the G8 specification, the treatment coefficients remain negative across all models, in line with the results from Table 3, although their magnitude declines once covariates are included. The interaction term for East Germany continues to be large and statistically significant, indicating that the fertility-reducing effect of the shortened school duration is again concentrated in the eastern states.

The interaction between the G8 treatment and university graduates is also negative across all models, suggesting that highly educated individuals experience an additional decline in fertility under the G8 regime. However, the estimates are relatively small and not significant. This muted pattern is consistent with the structure of the data. Because the G8 analysis includes Abitur cohorts from 2000 to 2015, many individuals who proceed to university are still relatively young within the observation window. Given that university graduates tend to postpone childbearing, a substantial fraction of their fertility lies beyond the observed period. As a result, the variation necessary to precisely estimate heterogeneous treatment effects among graduates is limited, leading to weaker statistical significance despite the consistent negative direction of the coefficient.

The time-reversed G9 specification provides a useful benchmark. When modeling the hypothetical extension of school duration by one year, the interaction term for university graduates becomes large and statistically significant, with coefficients ranging from 0.113 to 0.322. This suggests that, if anything, an additional year of schooling would increase fertility particularly among highly educated individuals. In this specification, graduates are effectively observed at older ages, which allows their fertility behavior to be captured more completely.

Taken together, the results from Table 5 indicate that both the realized G8 reform and the hypothetical G9 scenario reflect a common underlying mechanism. While differences in statistical significance stem from limitations in the observable fertility window rather than from genuine behavioral heterogeneity. When age and sample constraints are relaxed in the G9 specification, the fertility response of university graduates becomes clearly identifiable across both East and West Germany. The results indicate that reducing schooling time decreases fertility most strongly in East Germany, while the G9 counterfactual suggests additional effects among university graduates.

6. Discussion

The two empirical approaches examine the same policy intervention from opposite directions, both alter school duration while holding compulsory schooling constant. Although the coefficients differ in sign due to the direction of the reform, both designs yield the same substantive conclusion: school duration causally affects fertility behavior. Extending schooling increases fertility, particularly among university graduates, whereas reducing schooling lowers fertility. While our first approach is reliable and follows the conventions of general difference-in-differences approaches, we acknowledge that the robustness check estimation relies strongly on the assumption of homogeneous treatment effects to bypass the arguments of the "forbidden regression" and is therefore limited in its reliability of the identified effect.

Overall, however, the evidence implies that shortening upper secondary schooling reduces total fertility in the first approach and the completed fertility in the robustness check. Specifically, the implementation of the reform resulted in 13.3 fewer children born on average per 100 women over the studied period compared to before the reform. This effect is mainly driven by the population of Eastern Germany, where the implementation of the reform resulted in an additional 25.7 fewer children per 100 women. The robustness check using the time-inversed model exhibits the same directionality, although the magnitudes differ slightly due to differences in the underlying variables. Across the alternative model specifications, an additional year of schooling (G9) is associated with 16.9 to 19.5 more children per 100 women.

At first glance, these findings appear counterintuitive. Standard human capital theory and much of the international literature associate more education (in years and certificates) with delayed childbearing and often lower completed fertility, at least among women. Yet here, a shorter period in secondary school is linked to lower fertility. In this case, however, the same education is delivered in fewer years. In the following, we unpack this apparent paradox and discuss how our results can be understood in light of theories of human capital accumulation, opportunity costs, and the norms of role transition over the course of life.

Put differently, how can shorter schooling, at unchanged levels of formal attainment, reduce fertility? A useful starting point is the classical distinction in Blossfeld & Huinink (1991)

between enrollment effects and attainment or resource effects on family formation. They argue that schooling affects family formation through two conceptually different channels: The first channel concerns norms of role transition. While individuals are enrolled in school or training, they are socially defined as "not yet adults". Marriage and motherhood are normatively incompatible with student status, which strongly depresses the hazards of first marriage and first birth during enrollment.

The second channel operates after schooling is completed: higher levels of education and better career prospects raise the opportunity costs of time devoted to childrearing. For motherhood in particular, Blossfeld & Huinink (1991) find that women with greater career resources tend to postpone or even forego having children. In their West German cohorts, prolonged enrollment mainly delayed marriage, whereas human capital and career resources more directly shaped the timing and incidence of first births.

The G8 reform shifts the balance between these two mechanisms in a very specific way: it reduces the length of the student role in upper secondary education by one year, potentially weakening the normative barrier ("pupils don't have children") for that one year. At the same time, it brings forward the entire sequence of further transitions (Abitur, entry into university, entry into the labor market) without reducing the curriculum. Existing work suggests that G8 students are roughly ten months younger at graduation but achieve similar levels of human capital and academic achievement (Dörsam & Lauber, 2015; Huebener & Marcus, 2015). Thus, the reform has likely accelerated human capital accumulation and career entry, rather than reducing them.

Our findings suggest that, for the cohorts studied, the human-capital and opportunity-cost mechanism dominates the enrollment-norm mechanism. The one-year reduction in time spent in the "student" role does not translate into more births before age 35. Instead, earlier entry into tertiary education and the labor market appears to compress and intensify the early career phase, in which women accumulate valuable experience and wage growth (Mincer, 1974).

When these women reach ages at which family formation becomes more salient, their opportunity costs of childbearing in terms of forgone earnings, promotion prospects, and career continuity are higher than for comparable G9 cohorts. In other words, the reform likely increased the shadow price of time spent in motherhood at any given chronological age, especially for university-educated women. The "extra" year that is freed from schooling is not converted into a year of low-opportunity-cost family formation, but into an additional year of human capital accumulation. Under this interpretation, our results are fully consistent with G. S. Becker's (1960) and Joshi's (1990) view of children as costly normal goods, whose demand declines as the opportunity cost of parental time rises. Put differently: if we "hold all other life-course stages constant", a shorter schooling duration should extend the biological window available for childbearing. But in reality, the life course is re-optimized: women use the earlier end of schooling to start tertiary education earlier, enter the labor market earlier, and invest more intensively in their careers before considering motherhood. Biological time constraints are thus not simply relaxed by one additional year; the reform simultaneously intensifies economic constraints in that additional year.

A notable aspect of our results is the stronger fertility response in East Germany. This regional heterogeneity is consistent with prior evidence showing that the East retains higher female full-time employment (Goldstein & Kreyenfeld, 2011), and institutional arrangements more compatible with early and continuous maternal employment (O. A. Becker et al., 2010;

Dornseiff & Sackmann, 2003). Under such conditions, the compression of early life-course stages induced by G8 may have intensified opportunity costs more sharply than in the West, where part-time employment and more traditional gender norms remain prevalent. Conversely, in the G9 specification, the largest positive effects also appear in the East, suggesting that extended schooling aligns more favorably with institutional support for combining work and family. Overall, these patterns indicate that the fertility effects of schooling reforms are mediated not only by individual opportunity costs but also by region-specific institutional and normative contexts.

7. Conclusion

This paper shows that even a seemingly modest shift in the structure of schooling, namely reducing the Gymnasium track from nine to eight years, produced a measurable and economically meaningful change in family formation in Germany. Across two complementary identification strategies, the evidence is remarkably consistent. Shortening upper secondary schooling reduces fertility up to age 35, while extending schooling would have increased it. These effects appear to reflect genuine behavioral responses, strongest in East Germany and, in the robustness check, among university graduates, groups for whom opportunity costs and institutional environments interact most sharply with compressed educational trajectories.

The findings overturn the intuitive expectation that freeing up a year by shortening schooling should ease biological time constraints and encourage earlier motherhood. Instead, the reform accelerated transitions into tertiary education and the labor market, intensifying early career investment precisely in the years when fertility is most biologically feasible. The additional year gained from schooling compression did not expand the window for family formation. It shifted it into a phase of higher opportunity costs, thereby reducing fertility.

Taken together, the results call for rethinking how education reforms interact with family-related policy goals. The timing of life-course transitions can be just as consequential as the level of educational attainment. Reforms that reorganize when key stages of schooling and career entry occur can unintentionally reshape fertility patterns even when total years of education remain unchanged. As many countries revisit the design of upper secondary schooling, curriculum compression, or pathways into higher education, these findings suggest that such reforms carry demographic implications that reach far beyond the classroom.

Germany's G8 experiment ultimately demonstrates that the organization of schooling has the power to shape the rhythms of work, partnership, and parenthood. In the coming decade, as treated and untreated cohorts age into their forties, future research will be able to reassess whether the fertility effects identified here persist once completed fertility histories become observable. Follow-up studies will be essential for determining whether the patterns found in this paper represent permanent shifts in completed fertility or temporary postponements that may later be offset.

References:

Becker, G. S. (1960). An Economic Analysis of Fertility. In *Demographic and Economic Change in Developed Countries* (pp. 209–240). Columbia University Press. https://www.nber.org/books-and-chapters/demographic-and-economic-change-developed-countries/economic-analysis-fertility

Becker, G. S., & Lewis, H. G. (1973). On the Interaction between the Quantity and Quality of Children. *Journal of Political Economy*, *81*(2, Part 2), S279–S288. https://doi.org/10.1086/260166

Becker, O. A., Lois, D., & Nauck, B. (2010). Differences in Fertility Patterns between East and West German Women. Disentangling the roles of cultural background and of the transformation process. *Comparative Population Studies*, *35*(1). https://doi.org/10.12765/CPoS-2010-02

Black, S. E., Devereux, P. J., & Salvanes, K. G. (2008). Staying in the Classroom and Out of the Maternity Ward? The Effect of Compulsory Schooling Laws on Teenage Births. *The Economic Journal*, *118*(530), 1025–1054. https://doi.org/10.1111/j.1468-0297.2008.02159.x

Blossfeld, H.-P., & Huinink, J. (1991). Human Capital Investments or Norms of Role Transition? How Women's Schooling and Career Affect the Process of Family Formation. *American Journal of Sociology*, *97*(1), 143–168. https://doi.org/10.1086/229743

Cengiz, D., Dube, A., Lindner, A., & Zipperer, B. (2019). The Effect of Minimum Wages on Low-Wage Jobs*. *The Quarterly Journal of Economics*, *134*(3), 1405–1454. https://doi.org/10.1093/qje/qjz014

Cygan-Rehm, K., & Maeder, M. (2013). The effect of education on fertility: Evidence from a compulsory schooling reform. *Labour Economics, European Association of Labour Economists 24th Annual Conference, Bonn, Germany, 20-22 September 2012*, *25*, 35–48. https://doi.org/10.1016/j.labeco.2013.04.015

DeCicca, P., & Krashinsky, H. (2023). The effect of education on overall fertility. *Journal of Population Economics*, *36*(1), 471–503. https://doi.org/10.1007/s00148-022-00897-y

Dornseiff, J.-M., & Sackmann, R. (2003). Familien-, Erwerbs- und Fertilitätsdynamiken in Ost- und Westdeutschland. In W. Bien & J. H. Marbach (Eds.), *Partnerschaft und Familiengründung:*

Ergebnisse der dritten Welle des Familien-Survey (pp. 309–348). VS Verlag für Sozialwissenschaften. https://doi.org/10.1007/978-3-322-95055-0_11

Dörsam, M., & Lauber, V. (2015). *The Effect of a Compressed High School Curriculum on University Performance*. https://www.econstor.eu/handle/10419/140876

Fort, M., Schneeweis, N., & Winter-Ebmer, R. (2016). Is Education Always Reducing Fertility? Evidence from Compulsory Schooling Reforms. *The Economic Journal*, *126*(595), 1823–1855. https://doi.org/10.1111/ecoj.12394

Goldstein, J. R., & Kreyenfeld, M. (2011). Has East Germany Overtaken West Germany? Recent Trends in Order-Specific Fertility. *Population and Development Review*, *37*(3), 453–472. https://doi.org/10.1111/j.1728-4457.2011.00430.x

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

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

Huebener, M., & Marcus, J. (2015). G8 high school reform results in higher grade repetition rates and lower graduate age, but does not affect graduation rates. *DIW Economic Bulletin*, *5*(18), 247–255. https://www.econstor.eu/handle/10419/110198

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

Joshi, H. (1990). The Cash Opportunity Costs of Childbearing: An Approach To Estimation Using British Data. *Population Studies*, *44*(1), 41–60. https://doi.org/10.1080/0032472031000144376

Koebe, J., & Marcus, J. (2022). The Length of Schooling and the Timing of Family Formation. *CESifo Economic Studies*, *68*(1), 1–45. https://doi.org/10.1093/cesifo/ifab014

Kreyenfeld, M. (2003). Crisis or Adaptation – Reconsidered: A Comparison of East and West German Fertility Patterns in the First Six Years after the 'Wende'. *European Journal of Population / Revue Européenne de Démographie*, *19*(3), 303–329. https://doi.org/10.1023/A:1024992712815

Kreyenfeld, M. (2004). Fertility Decisions in the FRG and GDR: An Analysis with Data from the German Fertility and Family Survey. *Demographic Research*, *S3*, 275–318. https://doi.org/10.4054/DemRes.2004.S3.11

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

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), e201–e253. https://doi.org/10.1111/geer.12162

Mincer, J. (1974). *Schooling, Experience, and Earnings. Human Behavior & Social Institutions No. 2*. National Bureau of Economic Research, Inc.

Monstad, K., Propper, C., & Salvanes, K. G. (2008). Education and Fertility: Evidence from a Natural Experiment. *The Scandinavian Journal of Economics*, *110*(4), 827–852. https://doi.org/10.1111/j.1467-9442.2008.00563.x

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

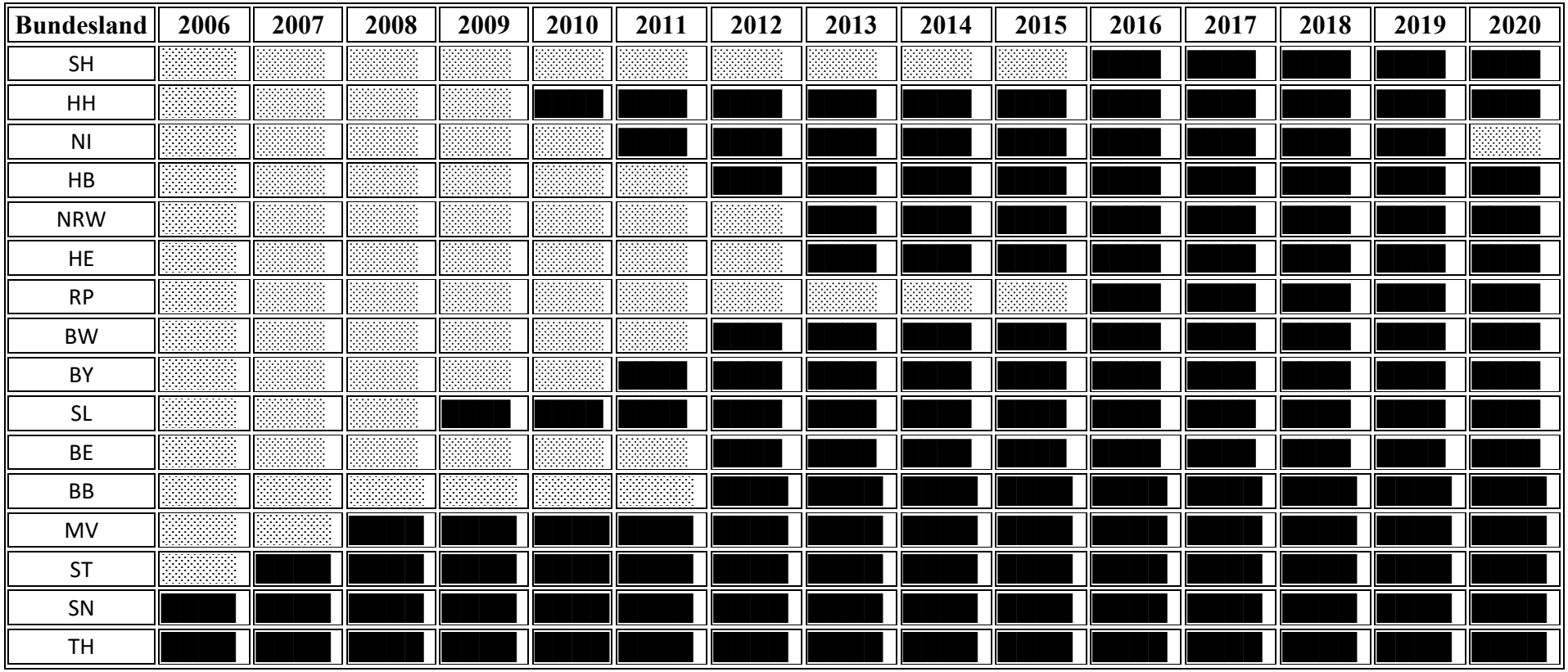
Silles, M. A. (2011). The effect of schooling on teenage childbearing: Evidence using changes in compulsory education laws. *Journal of Population Economics*, *24*(2), 761–777. https://doi.org/10.1007/s00148-010-0334-8

Vasireddy, S., Berrington, A., Kuang, B., & Kulu, H. (2023). Education and Fertility: A Review of Recent Research in Europe. *Comparative Population Studies*, *48*. https://doi.org/10.12765/CPoS-2023-21

Wang, G., Hamad, R., & White, J. S. (2024). Advances in Difference-in-differences Methods for Policy Evaluation Research. *Epidemiology*, *35*(5), 628. https://doi.org/10.1097/EDE.0000000000001755

Willis, R. J. (1973). A New Approach to the Economic Theory of Fertility Behavior. *Journal of Political Economy*, *81*(2, Part 2), S14–S64. https://doi.org/10.1086/260152

Table 1: G8 and G9 Abitur graduation cohorts by year and state



Notes: Dark cells indicate years in which the respective state had a G8 Abitur cohort graduating (i.e., the eight-year Gymnasium track was in force). Light gray/dotted cells indicate years in which the state had a G9 Abitur cohort graduating (nine-year track).
SH = Schleswig-Holstein, HH = Hamburg, NI = Niedersachsen, HB = Bremen, NRW = Nordrhein-Westfalen, HE = Hessen, RP = Rheinland-Pfalz, BW = Baden-Württemberg, BY = Bayern, SL = Saarland, BE = Berlin, BB = Brandenburg, MV = Mecklenburg-Vorpommern, ST = Sachsen-Anhalt, SN = Sachsen (always G8), TH = Thüringen (always G8).

Table 2: Summary Statistics

Summary statistics of the stacked DiD G8 implementation dataset. TFR is the dependent variable. G8_post_stack is the main treatment variable. Rel_year is the time variable.

The G8 Treatment					
Variable	Obs	Mean	Std. dev.	Min	Max
TFR	1,282	0.2330534	0.7042224	0	8.5
G8_post_stack	1,309	0.144385	0.351614	0	1
G8_east	1,309	0.0275019	0.1636032	0	1
G8_and_Uni	1,296	0.0408951	0.1981236	0	1
Net_income	861	1383.815	1003.974	26	9000
Publ_servant	824	0.336165	0.472683	0	1
Married	1,309	0.1390374	0.3461178	0	1
Apprenticeship	1,307	0.5967865	1.309205	0	6
stack_state	1,309	14.04431	6.868547	1	26
rel_year	1,299	-2.65204	5.078029	-14	8
rel_year_pos	1,299	12.34796	5.078029	1	23

Table 3: Stacked Difference-in-Differences

The following table presents the baseline model (Specification I) of the implementation of the G8 reform. Specification II adds an additional 'Treatment x East Germany' interaction to distinguish fertility effects between East and West Germany.

	Specification I			Specification II		
	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
Treatment (G8)	-0.275*** (0.0833)	-0.155** (0.0657)	-0.133** (0.0612)	-0.201** (0.0804)	-0.0909 (0.0654)	-0.0698 (0.0604)
Treatment (G8) x East	-	-	-	-0.321*** (0.0878)	-0.257** (0.0954)	-0.257*** (0.0916)
Net income	-	7.65e-06 (2.12e-05)	-7.44e-07 (2.05e-05)	-	6.25e-06 (2.11e-05)	-1.60e-06 (2.02e-05)
State Officer	-	0.000515 (0.0426)	0.00383 (0.0366)	-	-0.00121 (0.0421)	0.00118 (0.0362)
Married	-	1.145*** (0.126)	1.116*** (0.129)	-	1.143*** (0.125)	1.112*** (0.128)
Apprenticeship / Vocational Training	-	-	0.101** (0.0441)	-	-	0.0979** (0.0447)
Vocational School, Healthcare Sector	-	-	0.117* (0.0629)	-	-	0.119* (0.0633)
Healthcare Sector Schooling completed before 1999	-	-	0.486*** (0.115)	-	-	0.496*** (0.115)
Technical College Degree / Master Craftsman Qualification	-	-	0.196 (0.169)	-	-	0.187 (0.164)
Civil Service Training	-	-	0.0184 (0.0477)	-	-	0.0157 (0.0445)
Other Qualification	-	-	-0.0643 (0.154)	-	-	-0.0646 (0.153)
Stack State FE	Yes	Yes	Yes	Yes	Yes	Yes
Stack Time FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	1,272	803	801	1,272	803	801
R^2	0.102	0.527	0.535	0.105	0.530	0.538

Note: Standard errors clustered at the state level in parentheses. ***, **, * denote statistical significance at the 1%, 5% and 10% levels, respectively.

Table 4: Time-inversed stacked Difference-in-Differences

The following table presents the baseline model (Specification IV) of the time inversed hypothetical implementation of the G9 schooling system. Specification V adds an additional 'Treatment x East Germany' interaction to distinguish fertility effects between East and West Germany.

	Specification IV			Specification V		
	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
Treatment (G9)	0.169** (0.0674)	0.195* (0.101)	0.184* (0.0953)	0.118* (0.0595)	0.117 (0.0855)	0.108 (0.0872)
Treatment (G9) x East	-	-	-	0.231 (0.187)	0.321** (0.153)	0.319** (0.140)
Net income	-	5.59e-05 (3.46e-05)	4.87e-05 (3.50e-05)	-	5.75e-05 (3.44e-05)	5.06e-05 (3.46e-05)
State Officer	-	0.0121 (0.0290)	0.0120 (0.0234)	-	0.0134 (0.0288)	0.0137 (0.0231)
Married	-	0.676*** (0.121)	0.681*** (0.117)	-	0.668*** (0.117)	0.672*** (0.113)
Apprenticeship / Vocational Training	-	-	0.142*** (0.0357)	-	-	0.139*** (0.0348)
Vocational School, Healthcare Sector	-	-	-0.0682 (0.0422)	-	-	-0.0688 (0.0445)
Technical College Degree / Master Craftsman Qualification	-	-	0.337 (0.430)	-	-	0.314 (0.409)
Civil Service Training	-	-	0.00459 (0.0702)	-	-	1.28e-05 (0.0640)
Other Qualification	-	-	-0.137 (0.104)	-	-	-0.143 (0.0983)
Stack State FE	Yes	Yes	Yes	Yes	Yes	Yes
Stack Time FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	1,901	1,088	1,086	1,901	1,088	1,086
R^2	0.074	0.234	0.241	0.076	0.239	0.246

Note: Standard errors clustered at the state level in parentheses. ***, **, * denote statistical significance at the 1%, 5% and 10% levels, respectively.

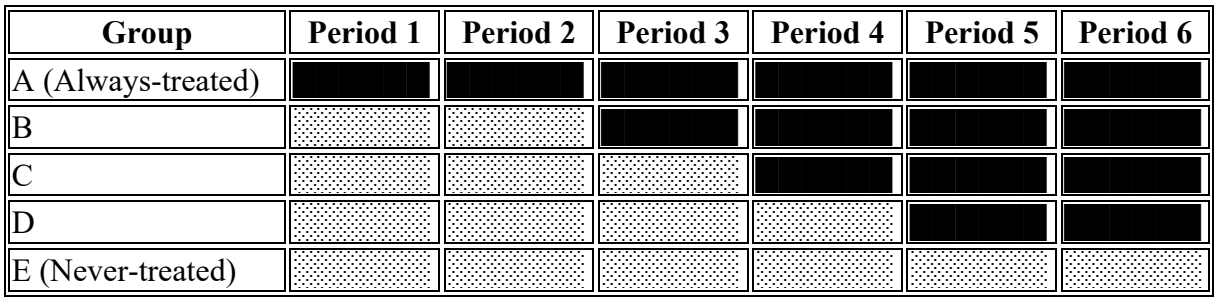
Table 5: Stacked Difference-in-Differences & Time-inversed stacked Difference-in-Differences

In the following table, Specification III extends the analysis from Table 3 (Specification II) by adding an additional 'Treatment x University Graduates' interaction to capture heterogeneity in fertility responses across educational groups. Specification VI adds the same for Table 4 (Specification V).

	Specification III			Specification VI		
	G8 Treatment			G9 Treatment		
	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
Treatment	-0.185**	-0.0635	-0.0564	-0.0139	0.0566	-0.00351
	(0.0849)	(0.0709)	(0.0646)	(0.0709)	(0.101)	(0.107)
Treatment x East	-0.324***	-0.262***	-0.259***	0.240	0.329**	0.333**
	(0.0878)	(0.0915)	(0.0888)	(0.192)	(0.152)	(0.136)
Treatment x Graduates	-0.0404	-0.0812	-0.0476	0.322***	0.113*	0.208***
	(0.0302)	(0.0486)	(0.0540)	(0.0577)	(0.0563)	(0.0651)
Net income	-	1.04e-05	1.69e-06	-	4.76e-05	2.94e-05
		(2.26e-05)	(2.18e-05)		(3.69e-05)	(3.95e-05)
State Officer	-	-0.000176	0.00249	-	0.0111	0.0112
		(0.0424)	(0.0363)		(0.0301)	(0.0235)
Married	-	1.142***	1.113***	-	0.660***	0.663***
		(0.126)	(0.128)		(0.117)	(0.110)
Apprenticeship / Vocational Training	-	-	0.0979**	-	-	0.175***
			(0.0462)			(0.0319)
Vocational School, Healthcare Sector	-	-	0.119*	-	-	-0.0518
			(0.0635)			(0.0465)
Healthcare Sector Schooling completed before 1999	-	-	0.489***	-	-	-
			(0.116)			
Technical College Degree / Master Craftsman Qualification	-	-	0.188	-	-	0.397
			(0.165)			(0.433)
Civil Service Training	-	-	0.0142	-	-	-0.00994
			(0.0447)			(0.0619)
Other Qualification	-	-	-0.0615	-	-	-0.149
			(0.153)			(0.0935)
Stack State FE	Yes	Yes	Yes	Yes	Yes	Yes
Stack Time FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	1,259	793	791	1,891	1,084	1,082
R^2	0.105	0.530	0.538	0.088	0.240	0.252

Note: Standard errors clustered at the state level in parentheses. ***, **, * denote statistical significance at the 1%, 5% and 10% levels, respectively.

Figure 1: Simplified Structure of Data Set



Notes: Black cells indicate periods in which a group is treated (G8), while grey cells indicate periods without treatment (G9).

Appendix:

Table A1: Summary Statistics

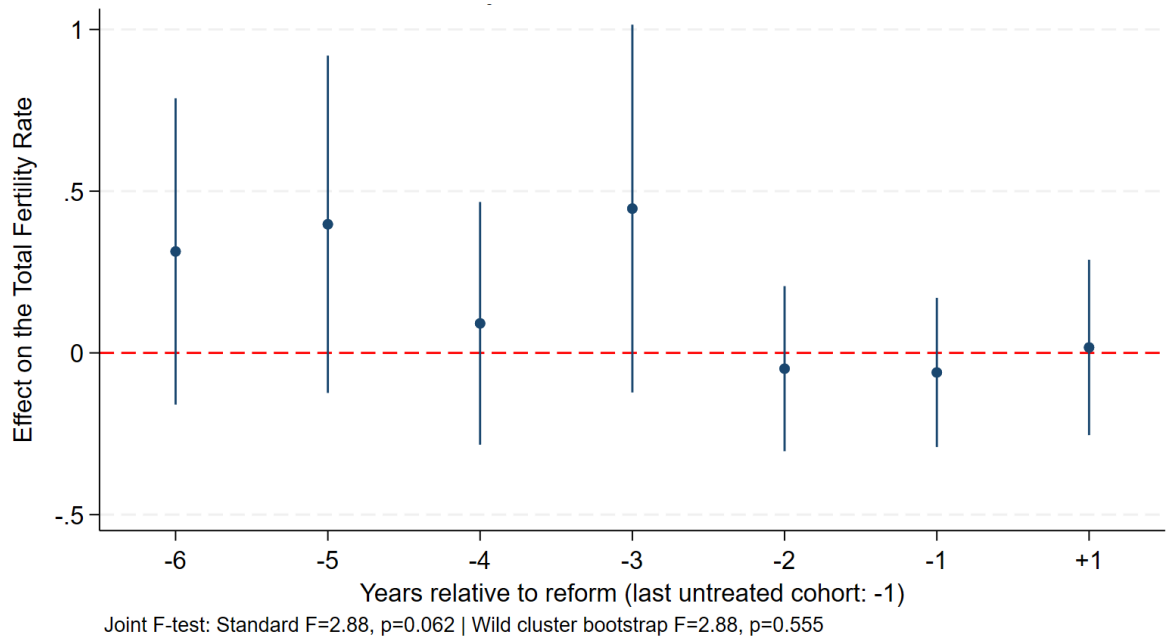
Summary statistics of the inverse stacked DiD hypothetical G9 implementation dataset. TFR is the dependent variable. G9_post_stack is the main treatment variable. Rel_year is the time variable.

The G9 Treatment					
Variable	Obs	Mean	Std. dev.	Min	Max
TFR	1,967	0.2064238	0.684141	0	8.5
G9_post_stack	1,978	0.2770475	0.4476533	0	1
G9_east	1,978	0.054095	0.2262624	0	1
G9_and_Uni	1,967	0.1052364	0.3069358	0	1
Net_income	1,186	1289.832	995.801	40	9000
Publ_servant	1,145	0.3484716	0.4766945	0	1
Married	1,978	0.1334681	0.340166	0	1
Apprenticeship	1,976	0.4473684	1.106145	0	6
stack_state	1,978	13.14712	7.619564	1	28
rel_year	1,910	-1.750262	6.664994	-17	12
rel_year_i_pos	1,910	13.75026	6.664994	0	29
rel_year_pos	1,910	16.24974	6.664994	1	30

Table A2: Common Trend Test

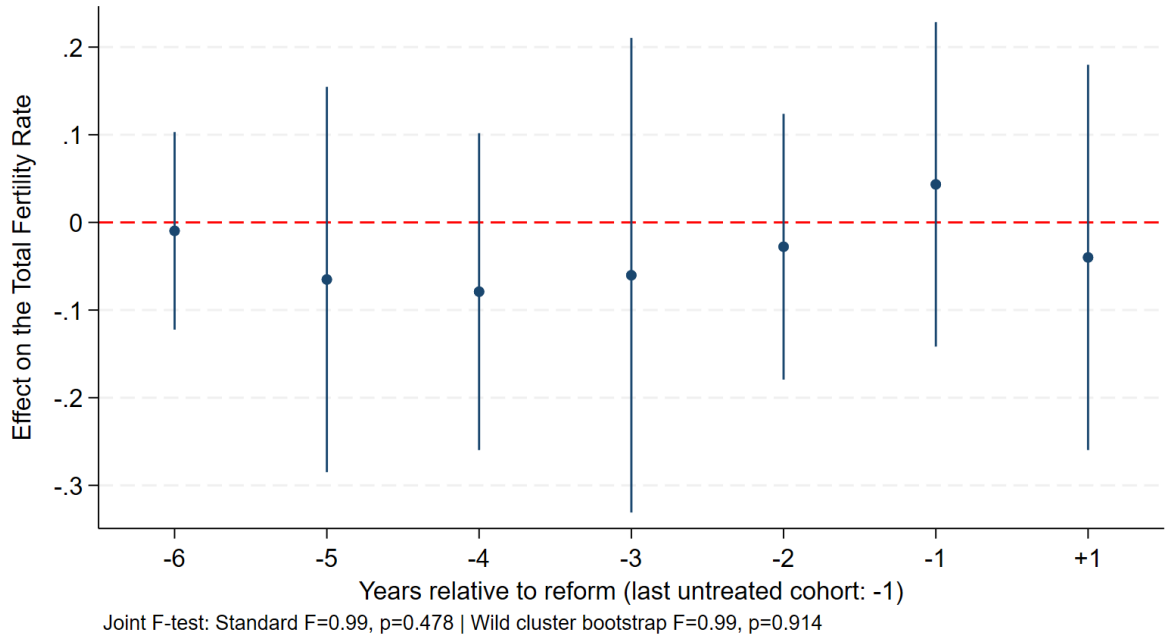
Common Trend Test		
Variables	(G8)	(G9)
Treatment/Control	0.173 (0.136)	-0.0526 (0.0558)
Constant	0.196 (0.116)	0.0919 (0.0513)
Stack State FE	Yes	Yes
Stack Time FE	Yes	Yes
Observations	890	834
R2	0.127	0.081

Figure A1 Event Study: Test for the Parallel Pre-Trends in the G8 Setting



Notes: This figure plots event-study coefficients from a stacked difference-in-differences regression of the total fertility rate on year-relative-to-reform bins interacted with treatment status, for the main G8 approach comparing states that implemented the G8 reform to states that continuously retained the 9-year secondary education track (G9). The regression includes state (stacked) fixed effects and birth-year fixed effects, with standard errors clustered at the state level. The reference period is the last untreated cohort (year -1); the sample is restricted to a balanced event window (-6 to +1) in which all reform cohorts are represented. Vertical bars denote 95% confidence intervals. A conventional joint F-test on the six pre-treatment event-study coefficients yields $F(6,11) = 2.88$, $p = 0.062$. Given the small number of clusters (12 states), we additionally report a wild cluster bootstrap test (Webb weights, 9,999 replications), which yields an identical F-statistic but a substantially higher p-value ($p = 0.555$), indicating that the conventional test's near-significance is an artifact of the small-cluster asymptotic approximation rather than evidence of a genuine pre-trend. We therefore conclude that the common trend assumption underlying our difference-in-differences design is valid, supporting the causal interpretation of our main estimates.

Figure A2 Event Study: Test for the Parallel Pre-Trends in the G9 Setting



Notes: This figure plots event-study coefficients from a stacked difference-in-differences regression of the total fertility rate on year-relative-to-reform bins interacted with treatment status, for the robustness check comparing states with a hypothetical G9 implementation to states that consistently maintained the 8-year secondary education track (G8). The regression includes state (stacked) fixed effects and birth-year fixed effects, with standard errors clustered at the state level. The reference period is the last untreated cohort (year -1); the reform-year cohort is excluded due to double graduation cohorts (Doppeljahrgänge) arising from the transition. Vertical bars denote 95% confidence intervals. A conventional joint F-test on the six pre-treatment event-study coefficients yields $F(6,11) = 0.99$, $p = 0.478$. Given the small number of clusters (12 states), we additionally report a wild cluster bootstrap test (Webb weights, 9,999 replications), which yields an identical F-statistic and confirms the absence of a pre-trend ($p = 0.914$). Together with the results in Figure A1, this supports the validity of the common trend assumption underlying our difference-in-differences design.