Working Paper No. 2017-04

The Impact of Schooling Intensity on Student Learning: Evidence from a Quasi-Experiment

Vincenzo Andrietti
University of Chieti-Pescara

Xuejuan Su
University of Alberta

Revised April 2018

The Impact of Schooling Intensity on Student Learning: Evidence from a Quasi-Experiment[*]

Vincenzo Andrietti[†]   Xuejuan Su[‡]

April 2018

Abstract

This paper uses a quasi-natural policy experiment in Germany, the G8 reform, to examine the impact of schooling intensity on student learning. The G8 reform compresses secondary school for academic-track students from nine to eight years, while holding fixed the overall academic content and total instruction time required for graduation, resulting in a higher schooling intensity per grade. Using German extension of the PISA data, we find that this reform improves test scores on average, but the effect differs across subgroups of students. The reform effect is larger for girls than for boys, for students with German born parents instead of immigrant parents, and for students having more books at home. The heterogeneous reform effects cannot be explained by changes in observed channels. Instead, quantile regression results suggest that unobserved heterogeneity plays an important role: while high-performing students significantly improve their test scores, the lowest-performing students hardly improve at all after the reform. We interpret the unobserved heterogeneity as reflecting students' capability to cope with the increase in schooling intensity.

JEL classification: I21, I28, D04

Keywords: Schooling intensity, learning outcome, G8 reform, heterogeneity, PISA.

[†]Università "G. d'Annunzio", Dipartimento di Scienze Filosofiche, Pedagogiche ed Economico-Quantitative (DiSFPEQ), Viale Pindaro 42, 65127 Pescara, Italy. E-mail: vincenzo.andrietti@unich.it.

[‡]University of Alberta, Department of Economics, 8-14 Tory Building, Edmonton, Alberta Canada T6G 2H4 E-mail: xuejuan1@ualberta.ca.

1 Introduction

Duration and intensity of schooling are two factors that jointly shape the curriculum and determine the effectiveness of the schooling process. The effect of schooling duration, measured as years of schooling, on student learning outcomes is relatively straightforward. The existing literature has repeatedly found that more years of schooling lead to better learning, measured by either labor market outcomes (see, for example, Ashenfelter and Kruger, 1994; Heckman and Vytlacil, 1998; Card, 2001; Heckman et al., 2006) or academic outcomes at the next education stage (Morin, 2013; Krashinsky, 2014). On the other hand, the effect of schooling intensity, i.e., the amount of academic content covered in a given year, can be hard to gauge. If too many topics are crammed in a year, students may struggle to keep up. If too few topics are covered, students may get bored. Thus, an appropriate level of intensity should adequately challenge but not overwhelm students. This focus of our paper is how schooling intensity affects student learning.

To empirically examine the impact of schooling intensity on learning outcomes, we need to overcome two major challenges. First, unlike schooling duration, measuring schooling intensity is difficult. Even if one has data on the amount of instruction time in a school year, it is typically hard to quantify the amount of academic content covered in those hours. When the additional instruction time is used to cover the same set of topics, schooling intensity does not change. This allows students to ask more questions, see more examples, and work on more practice problems, so they are all expected to do better even if the marginal returns are diminishing.[1] In contrast, when the additional instruction time is used to cover new

---

[1] A large literature studies the effect of instruction time on student achievement in this setting, exploiting variations induced by policies that lengthen the school day or the school year (Bellei, 2009; Parinduri, 2014; Kraft, 2015), shift state-mandated school start and/or test dates (Sims, 2008; Fitzpatrick et al., 2011; Agüero and Beleche, 2013; Carlsson et al., 2015; Aucejo and Romano, 2016), switch to four-day school weeks (Anderson and Walker, 2015), reallocate instruction time into a certain subject (Taylor, 2014; Cortes et al., 2015; Dougherty, 2015), or those induced by natural events leading to unscheduled school closings (Marcotte, 2007; Marcotte and Hemelt, 2008). Lavy (2015) and Rivkin and Schiman (2015) exploit within-student variation in subject-specific instruction time in the PISA data, using student fixed effects to control for unobserved heterogeneity. Consistent with most previous studies, they find that instruction time has significantly positive effects on test scores.

topics, schooling intensity increases. This becomes more demanding on students, especially those who are already struggling to keep up (Cuhna and Heckman, 2007; Clotfelter et al., 2015). So the same instruction time may correspond to different levels of intensity, depending on the breadth and depth of the topic coverage. Second, to the extent schooling intensity can be measured, the observed variations in the intensity levels are generally endogenous. Students may self-select into a suitable intensity level to improve their learning outcomes, and teachers may also adjust the intensity level to better serve their students. To avoid potential biases due to self-selection or endogeneity, we need to find a source of exogenous variation in schooling intensity.

In this paper, we overcome both challenges by using a quasi-natural policy experiment in Germany which we call the G8 reform. This reform compresses the duration of secondary schooling for academic-track students from nine (G9) to eight (G8) years, but holds the total amount of academic content and instruction time required for high school graduation fixed. As a result, each school year has more hours of instruction *and* covers more academic content, leading to an unambiguous increase in the intensity level. Furthermore, the G8 reform can be viewed as a quasi-experiment because it is driven by concerns over demographic changes and labor market conditions in Germany, instead of concerns over the schooling process *per se*. In particular, the aging population and the shortage of skilled labor are the main reasons why German states adopted the G8 reform, such that their college graduates can join the labor force one year earlier. In this sense, the G8 reform provides a source of exogenous variation in schooling intensity across German states and over time.

With this reform, learning outcomes can be measured when students are either in high school or at the end of high school. We prefer the former measure because learning outcomes at a certain mid-grade of high school reflect only the intensity effect, since schooling duration up to this mid-grade is the same as before. In contrast, learning outcomes at the end of high school would capture both impacts, one from shorter duration and the other from higher intensity. For our analysis, this mid-grade is grade nine, when students are assessed in the

German extension of the Programme for International Student Assessment (PISA).[2] Using five waves of PISA data, we find that higher schooling intensity has a significant and positive impact on the average scores of the three subjects tested: reading, mathematics, and science. We also find evidence of heterogeneous reform effects across students. In particular, the reform effect is larger for girls than for boys, for students with German born parents than for those with immigrant parents, and for students having more books at home.

To better understand the mechanisms through which the reform effects arise, we explore several possible channels to the extent we have data. First, by adding more instruction time in school, we consider whether the reform reduces time available for out-of-school activities, especially those that are academically productive such as homework, extracurricular programs, and/or remedial work for struggling students: If the higher schooling intensity "crowds out" other beneficial out-of-school activities, the net effect of the reform may depend on the trade-off between the two (Thile et al., 2014). Second, the reform could also affect teacher quality if it changes the composition of high school teachers. Finally, the reform could affect classroom quality if the higher intensity increases the stress level for teachers and/or students, leading to behavioral changes in the interactive classroom environment. We find no significant changes in these observed channels after the reform, either for the students population as a whole or across subgroups. Thus the reform effects do not appear to be driven by changes in these observed channels.

Lastly, we investigate whether the reform effects could be explained by unobserved heterogeneity. Using quantile regression, we estimate the reform effects at different quantiles of the test score distribution. After controlling for observed heterogeneity, we find that high-performing students (those at upper quantiles) have acquired more knowledge by ninth grade following the reform, while low-performing students (those at lower quantiles) have not. Thus,

---

[2]Other studies examine the G8 reform effects on student personality (Dahmann and Anger, 2014), cognitive skills at high school graduation (Dahmann, 2017), high school graduation rate and graduation age (Huebener and Marcus, 2017), and post-secondary enrollment (Meyer and Thomsen, 2016). We are the first to examine learning outcomes in high school, thus separating the intensity effect from the duration effect. Huebener et al. (2017) perform analysis similar to ours in Andrietti (2015, 2016) and in Andrietti and Su (2016).

unobserved heterogeneity that distinguishes high-performing from low-performing students does play a critical role in understanding the reform effects. We interpret this unobserved heterogeneity as reflecting students' capability to more or less effectively cope with the more demanding schooling process after the G8 reform.

The rest of the paper is structured as follows. Section 2 describes the G8 reform and PISA data used for our empirical analysis. Section 3 presents our regression models and discusses the identification strategy. Estimation results are reported in Section 4. Section 5 draws the conclusion.

2   Data

2.1   The G8 reform

In Germany, the sixteen federal states have jurisdiction over public education policies. Typically students spend four years in primary schooling, and then enter secondary schooling at the beginning of grade five.[3] Upon entering secondary schooling, students are "tracked" into three types of schools: the basic track, the middle track, and the academic track (also called *Gymnasium*). The former two types provide vocationally oriented schooling through grade nine or ten, while the latter prepares students for university entrance qualification called *Abitur*.[4] Our focus is on the academic-track students.

Before 2001, in all but two German states, the academic track lasted nine years, resulting in a total of thirteen years of schooling up to high school graduation.[5] Driven by concerns over demographic changes and labor market conditions, the fourteen states started to implement the G8 reform since 2001, some states earlier than others. While most states began the reform

---

[3]As an exception, secondary schooling begins in grade seven in three states: Berlin, Brandenburg, and Mecklenburg-Vorpommern (since 2007). On the other hand, Bremen and Lower-Saxony used to have secondary schooling starting in grade seven, but changed that to grade five in 2004 and 2003 respectively.

[4]A few states also have comprehensive schools that offer multiple tracks within the same school.

[5]The two exceptions are former Eastern German states Saxony and Thuringia, where the academic track lasts eight years.

on the entering student cohort, a few states extended the applicability to student cohorts already enrolled in the academic track. Details about the timing and the implementation of the reform by state can be found in Figure A.1 in the Appendix.

The G8 reform, as discussed earlier, reduces the academic track duration from nine to eight years, but holds the overall academic content and the total instruction time unchanged. More specifically, just like the G9 regime, a total of 265 year-week hours is required for high school graduation. The 265 year-week hours is the sum of the number of instruction hours per week across all academic-track grades, i.e., nine grades under G9 but eight grades under G8. As a result, the number of weekly instruction hours per grade increases after a state implements the G8 reform. These additional hours are used to cover new academic content that was previously taught in higher grades. Thus, the increase in instruction time due to the G8 reform directly translates into an increase in schooling intensity, because more academic content is covered in each of the school years.

2.2 PISA data

We use five waves (2000, 2003, 2006, 2009, and 2012) of the German extension of the Programme for International Student Assessment (PISA) data.[6] While the international version of the PISA data assess 15-year-old students, the German extension of the first three waves (PISA-E 2000, 2003, and 2006) enlarges the original PISA samples by collecting additional grade-9 and age-15 samples. In 2009 and 2012, (smaller) grade-9 samples are also collected in addition to the original age-15 samples. Because the German state information is not released in the age-15 sample of the PISA 2009 wave, for our analysis, we pool all grade-9 samples from the five waves, namely PISA-E 2000, 2003, and 2006, as well as PISA 2009 and 2012.

In each PISA cycle, a range of relevant skills and competencies are assessed in three subjects: Reading, mathematics, and science. Each subject is tested using a broad sample

---

[6]Baumert (2009); Klieme (2013); Prenzel (2007, 2010, 2015)

of tasks with differing levels of difficulty, so that the test score represents a coherent and comprehensive measure of student proficiency in the given subject.[7] Using item response theory, PISA maps student achievement in each subject on a standardized scale with a mean of 500 and a standard deviation of 100 across all OECD countries included in the study. Over the five waves, our samples include about 34 thousand observations in reading, and about 30 thousand observations in mathematics and science.[8]

Besides test scores, the PISA questionnaire also collects other background information from students, their parents, teachers, and school principals. For our analysis, we extract two sets of control variables, representing characteristics of the students and their schools respectively. Student controls include demographics, their family background, and socio-economic characteristics. School controls include school type, the size and the composition of its student body, and measures of school resources. Summary statistics on these variables are reported in Table 1.

Moreover, PISA also contains some information on student time usage both in school and out of school. In particular, PISA 2003 questionnaire to students includes the following question: "In the last full week you were in school, how many instruction hours (*each 45 minutes*) did you have *in total*?" PISA 2009 and 2012 questionnaire asks "In a normal, full week at school, how many instruction hours (*each 45 minutes*) do you have *in total*?" Although this self-reported information may be prone to measurement error, it nonetheless provides some corroborating evidence to the increase in weekly instruction hours as a result of the G8 reform. Accordingly, we construct a PISA instruction time variable by averaging self-reported weekly instruction hours at the school level. Comparing this variable for

---

[7]PISA test scores are averages of five plausible values, which are drawn from a distribution of values that a student with the given number of correct answers could achieve as a test score (OECD, 2012).

[8]The sample size difference is due to the fact that just over half of the students who took the reading test also took the standard PISA math and science tests in PISA-E 2000. Although supplementary national tests on math and science were given in a second day to all students who took the reading test, these national tests are different from the standard PISA test and more closely related to German curricula. To ensure comparability, our samples include only those students assessed by the standard PISA tests and excludes those assessed by the national tests, using the subject-specific final student weights available for PISA-E 2000.

both academic-track students and basic- and middle-track students allows us to investigate whether there are other unobserved common changes contemporaneous to the G8 reform.

3 Regression models

3.1 Difference-in-difference model

The staggered implementation of the G8 reform over time and across states enables us to use a difference-in-difference (DiD) model for identification:

$$zscore_{ist} = \beta \cdot G8_{st} + \alpha \cdot \mathbf{X_{ist}} + \delta_s + \gamma_t + \varepsilon_{ist}, \tag{1}$$

where $zscore_{ist}$ is the standardized test score for an academic-track student i in state s and year t. $G8_{st}$ is the indicator variable for the G8 reform status, which equals one if state s in year t has a student cohort treated by the G8 reform, and zero otherwise. X_{ist} is a vector of student and school controls. δ_s and γ_t are state and year fixed effects, and ε_{ist} is the residual error term. Our main interest is the coefficient β measuring the average effect the reform has on all academic-track students.

We employ three treatment definitions in the analysis. First, there is the G8 dummy described above. Students cohorts coded as treated are typically those subject to the G8 reform upon entering the academic track, except in the few states where the reform affected also cohorts already enrolled in the academic track.[9] We define the G8 indicator variable accordingly. Figure 1 displays the treatment status of each PISA cohort across the sixteen German states.

Second, the length of treatment may vary across states and/or cohorts. For example, there are cohorts treated in states where tracking takes place in grade seven (namely Berlin

---

[9]The state of Hesse introduced G8 in 2004 only in 10% of academic track schools. Given the low probability of treatment assignment, we code Hesse cohort in the 2009 PISA grade-9 sample (i.e., those entering the academic track in 2004) as control instead of treated. Excluding this specific cohort (or this state) does not affect the results qualitatively.

and Brandenburg in PISA 2009 and 2012, and Mecklenburg-Vorpommern in PISA 2012), and cohorts beginning treatment in grade seven or eight (i.e., Saxony-Anhalt and Mecklenburg-Vorpommern in PISA 2006). For these cohorts, the length of treatment until grade nine is shorter than the modal treatment duration of five years. To capture this heterogeneity, we also create a variable called "years of treatment".

Finally, we use official historical timetables (KMK, 1997-2014b) to compute the year-week hours of instruction, averaged over grades five through nine by state and cohort. This explicitly represents the schooling intensity before and after the G8 reform, which is state and cohort specific. When this variable is used in place of the G8 dummy in (1), the coefficient gives us the impact of schooling intensity increase by one additional instruction hour delivered over grades five to nine.

3.2 Heterogeneous reform effects

Our next interest is on the potential heterogeneity of the reform effects across students. More specifically, here we focus on observed heterogeneity across students. We augment the DiD model in (1) by allowing different reform effects for different subgroups of students:

$$zscore_{ist} = \Sigma_{g=1}^{N}\beta_g \cdot G8_{st} \cdot I(\text{student } i \in \text{ subgroup } g) + \alpha\mathbf{X_{ist}} + \delta_s + \gamma_t + \varepsilon_{ist}, \qquad (2)$$

where $I(\text{student } i \in \text{subgroup } g)$ is the indicator variable that equals one if student i belongs in one of N subgroups indexed by g. Our interest is the relationship of the coefficients β_g, each measuring a reform effect for a particular subgroup defined by a given student characteristic.

Last, besides observed heterogeneity, we are also interested in potential heterogeneity of the reform effects based on unobserved student characteristics. This is done by applying the DiD method in a quantile regression model:

$$h_{\tau,st} = \beta_\tau G8_{st} + \alpha_\tau \mathbf{X_{ist}} + \delta_{\tau,s} + \gamma_{\tau,t}, \qquad (3)$$

8

where $h_{\tau,st}$ is the test score at the τ-th quantile of the distribution in state s and year t. Since $\mathbf{X_{ist}}$ controls for observed heterogeneity, the quantile-specific coefficients β_τ thus represent reform effects on students at different quantiles of the test score distribution according to their unobserved heterogeneity.[10]

4   Results

4.1   Reform impact on instruction time

Table 2 reports how the G8 reform affects the weekly hours of instruction across grades. The reported coefficients are estimated using the baseline specification, which includes only the policy dummy together with state and year fixed effects. Columns 1 to 6 report the estimates based on the KMK official timetables. In contrast, columns 7 and 8 report the estimates based on the self-reported grade-9 weekly instruction hours collected in the PISA questionnaire on academic-track students and basic- and middle-track students, respectively.

We find that the G8 reform increases the average weekly instruction time by 2.5 hours over grades 5–9 (column 1), significant at the 1% level. However, the increase is by no means uniform across the grades. It is smaller in grades 5–6 and substantially bigger in grades 7–9, reflecting a phase-in of the increase in schooling intensity (columns 2-6). Using self-reported data, for academic-track students, the increase in weekly instruction hours is in line with that estimated using the official timetables. The self-reported increase is 2.3 hours, significant at the 1% level. In comparison, there is no significant change in the self-reported instruction time in lower tracks. This result lends support to the idea that the G8 reform is the source of the increase in instruction time for academic-track students, and rules out unobserved factors that are contemporaneous to the G8 reform and affect the instruction time for students in

---

[10]A limitation of the quantile regression is that, despite the importance of clustering standard errors at the treatment (state) level to avoid overstating precision (Bertrand et al., 2004) is widely recognized, a statistically valid method to cluster standard errors has not been developed yet. This is further complicated by the sampling weights associated with the observations in the complex survey design. As a result, we can only report the standard error assuming i.i.d. errors.

all tracks.

4.2 Average reform effect

Table 3 reports the average effect of the G8 reform on test scores, panel A for reading, panel B for math, and panel C for science. Each panel row reports coefficients (and standard errors) estimated from the DiD model using one of the three treatment variable: the G8 reform indicator, years of treatment, and the average weekly instruction hours over grades 5–9 according to the official timetables. The reported coefficients are obtained estimating equation (1) under different specifications with OLS, using PISA student sampling weights and considering the five plausible values per subject available for each student.[11] Standard errors are clustered on the state level to account for serial error correlation within each state over time.[12]

The results of the baseline specification are reported in column 1, those with student controls are reported in column 2, and finally those with both student and school controls are reported in column 3.[13] Overall, the estimates of the reform effects remain stable across these specifications. This implicitly corroborates the G8 reform as a quasi-experiment, since student and school characteristics do not appear to be correlated with the reform status, and their omissions do not change the estimates significantly. For improved estimation precision, we focus on the main specification (column 3) hereinafter.

We find that the G8 reform has a significantly positive effect on test scores. The reform

---

[11]Plausible values are generated as random draws from the distribution of scores that could be reasonably assigned to each student, accounting for missing information on questions outside the different subsets that students answer from the total item pool. We account for this multiple imputation procedure following the recommendations in OECD (2012): For each subject, OLS regressions are run separately on each of the five plausible values (standardized to have mean zero and variance in each PISA assessment). Results are then aggregated to obtain the final estimated coefficients and their respective standard errors using the STATA `pv` command.

[12]Although this approach may lead to over-rejection of the null hypotheses when the number of clusters (n) is small (Cameron and Miller, 2015), this does not appear to be an issue in our case ($n = 16$). The wild cluster bootstrap procedure (Cameron et al., 2008) produces qualitatively similar results.

[13]We use missing-value dummies to account for observations with missing values in the control variables, which represent less than 5% of all observations in the pooled sample. Results are similar if we drop the observations with missing values.

increases the average test score in reading by 0.078 standard deviations, in math 0.067, and in science 0.085. Each year of treatment leads to an increase of $0.013-0.015$ standard deviations, while an additional hour in weekly instruction time increases test scores by $0.022-0.034$ standard deviations. So using the modal treatment duration of five years (grades 5–9), or the average 2.5 hours increase of weekly instruction time from Table 2, the magnitudes of the reform effects are comparable in all three cases.[14]

The validity of our DiD results relies on a number of conditions, namely treated and control states should follow common trends in the absence of the reform, the reform should not induce significant compositional changes in the student body, and other contemporaneous reforms should not have a differential impact on students across treated and control states. Below we address these concerns.

First, regarding the common trend assumption, Figure 2 represents the inter-temporal reform effects estimated using the baseline specification. These effects are captured by breaking the G8 reform dummy into a set of indicator variables, one for the first treated cohort, two lead variables (three-year prior and six-year prior), and two lagged variables (three-year after and six-or-more-year after). The omitted category is the cohort nine-or-more-year prior to the first treated cohort in the PISA data. The pattern of the inter-temporal effects is consistent with the common trend assumption. In particular, the coefficients for the two lead dummies are statistically insignificant, suggesting that prior to the G8 reform there are no significant differences between students in treated and control states. Similarly, with regard to compositional changes, we find no evidence that the reform has any significant impact on observed student and school characteristics (see Tables A.1 and A.2 in the Appendix).

Next, we perform a number of robustness checks in Table 4. We consider a placebo

---

[14]To put this magnitude in context, we also estimate the duration effect using the age-15 sample in PISA-E 2000 and 2003 (before the G8 reform). Depending on the subject, the impact of an additional year of schooling (from grade nine to grade ten) is $0.47-0.56$ standard deviations. Note that the additional instruction hours over grades 5–9 after the G8 reform would correspond to approximately 40% of a typical school year before the reform, thus increasing the test scores by $0.19-0.22$ standard deviations. In this sense, the intensity effect is small compared to the duration effect. However, we caution that the estimated duration effect should not be interpreted as causal.

reform dummy which equals 1 for the cohort immediately before the first cohort actually treated by the G8 reform (column 2), where a significant estimate would indicate different outcomes for the treated and control groups *before* the G8 reform was implemented. We also consider the reform impact on basic and middle-track students, who should not be affected by this reform (column 3), where a significant estimate would indicate that there are other unobserved factors that affect these lower-track students and are contemporaneous to the G8 reform. Both placebo tests yield insignificant results. When we use the middle-track students as an additional control group in a DDD setting (column 4), the results are similar to those found using the DiD method. Similarly, when we allow treated and control states to follow different linear trends, the results remain robust (column 5).

The common trend assumption could also be violated as a consequence of contemporaneous education policy changes. A contemporaneous education reform that directly affects academic-track students is the introduction of Centralized Exit Examinations (CEEs). We add a dummy indicating the state-specific cohorts affected by the introduction of CEEs and re-estimate equation (1) (column 6), and the main results remain robust. Another contemporaneous change took place between 2003 and 2009, when a federal investment program promoted the introduction of all-day schooling in Germany. All-day schooling extends the school day by offering a school lunch program, extra-curricular activities, and supervised study hours, and it is typically attended on a voluntary basis. Again, controlling for the share of all-day students in a state (KMK, 2002-2012) does not change the main results (column 7). Finally, the first cohort treated by the G8 reform in a state is part of a *double graduating cohort*, since it graduates at the same time as the last cohort under the old G9 regime. This implies stronger competitive pressure on the students for university admissions and later job opportunities, making the first treated cohort different from subsequent cohorts. We include a dummy for the double cohort, and the main results remain unchanged (column 8).

4.3 Heterogeneous reform effects

Not all students can effectively cope with the higher schooling intensity, so the reform effects may be heterogeneous across students depending on their characteristics. We estimate (2) by dividing students into different subgroups. Table 5 reports the heterogeneous reform effects on six characteristics. The demographic variables include gender, age, and parental immigration status (proxied by whether parents were born abroad or in Germany). The socioeconomic background variables include parental education level, parental socioeconomic index (ISEI), and parental investment in child's human capital (proxied by the number of books at home). Results are reported separately for the three subjects.

First, girls exhibit earlier developmental patterns than boys in adolescence, in particular non-cognitive skills such as self-control and perseverance (Spinath et al., 2014), so they may better cope with the increase in schooling intensity. This is indeed what we find, namely in reading and science, girls have acquired more knowledge by ninth grade following the reform than boys. Along similar developmental consideration, we divide students into the very young (age in the lowest tercile) and the rest (age in the middle and upper terciles). We find that older students have acquired more knowledge by ninth grade following the reform than the very young ones, and the difference is statistically significant in math. We also find that students with foreign born parents (proxy for immigrants) see no change in their test scores following the reform, while those with German born parents improve significantly.

Next, regarding socioeconomic background, there is no significant difference according to parental education level (less than tertiary vs. tertiary and above). When we divide students by their parental ISEI (the lowest tercile versus the middle and upper terciles), we find some evidence that students do better in math if their parents have middle to high ISEI, even though the difference is only marginally significant at 10% level. Finally, regarding parental investment in child's human capital (proxied by the number of books at home), students with fewer books at home hardly improve their test scores at all after the reform, while those with more books at home improve significantly.

Overall, there is evidence that reform effects are indeed heterogeneous across students: it is generally larger for girls than for boys, for students with German born parents instead of immigrant parents, and for students having more books at home.[15] Such heterogeneous effects may raise concerns about the longer-term impact of the reform, namely how students would perform at high school graduation or beyond. For example, consider students with foreign born parents and/or having fewer books at home, for whom the intensity effect of the reform is essentially zero. With one less year of schooling, these students would arrive at high school graduation one year earlier but with less human capital, so they may be at risk of not meeting the graduation requirements. In these instances, grade repetition can serve as a useful tool to counter the negative duration effect, selectively applied to these at-risk students. This is indeed what the literature finds (Huebener and Marcus, 2017), namely the G8 reform significantly increases the probability of repeating a grade in the final three years before graduation.

4.4   Mechanisms

To better understand the mechanisms of the G8 reform effects, we examine several possible channels, to the extent we have data. For each of the potential channels, the reform could have potential impact on multiple outcomes reflecting the same main theme. We follow Deming (2009) and use these outcomes to construct a summary index, which has the advantage of reducing measurement error and is robust to the multiple inference problem. There are three steps involved in creating the summary index. First, each outcome is normalized to have mean zero and standard deviation of one. Second, a single index is created by averaging the relevant outcomes. Finally, the index itself is again normalized to have mean zero and standard deviation of one. So we have one summary index for each of the channels under consideration. The results are reported in Table 6.

The first channel we consider is student *time usage* for out-of-school activities, which

---

[15]More flexible, but less parsimonious, specifications allowing further interactions between the subgroup indicator and the other variables included in the models lead to similar results.

may be crowded out as the G8 reform increases in-school instruction time. In PISA 2003 and 2012 questionnaire, students were asked different questions about the amount of time (hours per week) they dedicated to self study (including homework and other forms), out-of-school classes (delivered off campus), and private tutoring. These activities are academically productive and complement classroom instruction to help students achieve better learning outcomes. Using the summary index for out-of-school study time (column 1), we find no evidence that the G8 reform had any significant impact, either for the student population as a whole, or across different subgroups (estimates not reported in the table). The second and related channel is student *probability* of attending classes beyond regular school hours. In PISA 2003 and 2009, students were asked different questions regarding whether they attended remedial classes (any subject or math in particular), and/or enrichment classes (any subject or math). These out-of-school classes are delivered on campus but beyond regular school hours, so they may be similarly crowded out as the G8 reform increases in-school instruction time. Furthermore, we expect that low-performing students would benefit more from remedial classes, while high-achieving students more from enrichment classes. Using the summary index for these class attendance probabilities (column 2), again we find no evidence that the G8 reform had any significant impact, either for the student population as a whole, or across different subgroups (estimates not reported). Collectively, these two channels suggest that the observed reform effects are not driven by changes in student behaviors out of school.

The third channel we consider is through teacher quality, which may change if the G8 reform affects the size and the composition of the teacher body. In all five waves (2000–2012) of PISA questionnaire, school principals were asked different questions about the size and the composition of the teacher body, including total number of teachers hired, the number of full-time versus part-time teachers, and the number of certified teachers. Information on the size of the teaching body and on the staff expenditures per student is also collected at the state level by the KMK (KMK, 2014a) and by the Statistische Bundesamt (SB) (Statistisches Bundesamt, 2015). Using the summary index on teacher quality (column 3),

we find no evidence that the G8 reform has any significant impact.[16]

Finally, the reform could have an impact on the classroom environment, including both student-teacher interactions and student-student interactions. In PISA 2000, 2003, 2009, and 2012, school principals were asked about a wide of issues that could hinder instruction or student learning, such as teacher absenteeism, teacher having low expectations of students, student disrupting class, student using drugs, student absenteeism, etc. Using the summary index on classroom quality (column 4), again we find no evidence that the reform has any significant impact. Overall, the reform effects do not appear to be driven by changes in these observed channels (out-of-school activities for students, teacher quality, or classroom environment), suggesting that unobserved factors may play an important role.

4.5   Quantile results

Since the observed channels exhibit little change after the reform, we now turn our attention to a channel through unobserved heterogeneity using quantile regressions (3). In the quantile regression, after controlling for observed heterogeneity, higher test scores can be interpreted as reflecting better learning capability (albeit unobserved or residue), and similarly lower test scores reflecting worse learning capability. The quantile regression results thus directly link the heterogeneous reform effects to the unobserved learning capacity at different quantiles of the test score distribution, without relying on student demographics and/or their parental background as a proxy.

The results are reported in Table 7. We see that after controlling for observed heterogeneity, low-performing students (those in the lowest deciles) do not benefit significantly from the reform, while high-performing students (those in upper deciles) benefit significantly. Furthermore, the magnitude of the reform effects generally increases as we move up the deciles, from 0.054 at the second decile in reading (the first significant result) to 0.104 at

---

[16]This empirical evidence is also consistent with anecdotal/document evidence OECD (2011), namely teachers union agreed to the extended school day without a comparable pay increase in exchange for not cutting teacher positions after the elimination of the last grade. The number of instruction hours per teacher remained essentially unchanged after the reform, only reallocated from the last to earlier grades.

the ninth decile, in math from 0.062 at the fourth decile to 0.084 at the ninth decile, and in science from 0.071 at the second decile to 0.101 at the ninth decile. Thus, the unobserved (or residual) learning capability that leads students to perform well in the old G9 regime also helps students further improve their performance after the G8 reform.

Given that low-performing students hardly improve at all (by grade 9) after the reform, just like those with foreign born parents and/or fewer books at home, they will be similarly at risk of not meeting the high school graduation requirement. Thus, based on both observed and unobserved heterogeneity, we find that the G8 reform widens the performance gap between struggling students and their peers.

5 Conclusion

Schooling intensity is a conceptually intuitive, and yet empirically under-explored, determinant of the effectiveness of the schooling process. We take advantage of a unique policy experiment, the G8 reform in Germany, to estimate the impact of an increase in schooling intensity on student achievement. Using PISA 2000–2012 data, we find that the increase in schooling intensity significantly improves test scores on average. However, the reform effect is small relative to that arising from a comparable increase in schooling duration, and furthermore, it is heterogeneous across students. Along observed dimensions, the reform effect is generally larger for girls than for boys, for students with German born parents instead of immigrant parents, and for students having more books at home. Along the unobserved dimension, high-performing students have acquired more knowledge by ninth grade following the reform, while this is not the case for low-performing students. Overall, the reform effects seem to depend critically on student capability to effectively cope with the schooling process, which becomes more demanding after the reform.

Our findings offer important policy insights related to the G8 reform. A major issue of the public debate over the G8 reform is whether it is possible to improve student learning

17

by increasing the intensity of schooling. Some fear the reform would overburden students and hence negatively affect their learning outcomes. In fact, some states are considering, or have already implemented, a (partial) switch back to the old G9 regime. Our results offer a two-prong answer to this debate. First, the average reform effect is positive, and the majority of the students appear capable of adapting to and indeed benefiting from the increase in schooling intensity. Second, there are indeed struggling students who would fall further behind compared to their peers, potentially resulting in worsened inequality. For these struggling students, well-targeted support measures (including grade repetition) will help alleviate the negative effects of the reform.

The implications of our paper go beyond the German context and shed light on the impact of related education reforms in other jurisdictions. For example in Canada, Ontario adopted a province-wide reform that shortened the high school duration from five to four years. However, unlike the G8 reform, the Ontario reform also cut the amount of required academic content in proportion to the reduction of schooling duration, thus leaving schooling intensity essentially unchanged. Two studies (Krashinsky, 2014; Morin, 2013) have found that after the reform, four-year high school graduates perform worse than their five-year counterparts in college. Our analysis can help decompose this total effect into two components, where the intensity effect is essentially zero (no change in schooling intensity) while the duration effect is negative (shorter duration). In Italy, The Ministry of Education is currently experimenting a pilot program similar to the G8 reform in Germany, i.e., shortening upper secondary education from five to four years while holding the academic requirement for graduation fixed.[17] Our results shed light on the expected impact of this reform on student learning, especially when at a mid-grade before high school graduation.

Beyond state sanctioned education reforms, our analysis also has broader implications. In general, schooling intensity may change as a result of personal choices at more disaggregate

---

[17]http://www.corriere.it/scuola/17_agosto_07/scuola-diploma-quattro-anni-via-sperimentazione-100-classi-ee746c20-7b56-11e7-8e8c-39c623892090.shtml, last accessed on October 18, 2017.

levels. For example, in the U.S., an increasing fraction of college students are non-traditional, part-time students, who experience a lower schooling intensity compared to full-time students. Our results would suggest that depending on their learning capability, this lower intensity may be beneficial if they would otherwise struggle to keep up with the full-time process, even though this comes at the expense of prolonging the schooling duration. Another example is the establishment of a magnet school in a public school district. If the magnet school puts more emphasis on certain academic subjects, its students will experience a higher schooling intensity in these areas. Our method provides a general framework to understand how this higher intensity would affect student learning, and evaluate potential mean-variance tradeoffs (average effect versus heterogeneous effects) in such situations.

References

Agüero, Jorge M., and Teresa Beleche (2013) 'Test-mex: Estimating the effects of school year length on student performance in mexico.' *Journal of Development Economics* 103, 353–316

Anderson, D. Mark, and Mary Beth Walker (2015) 'Does shortening the school week impact student performance? evidence from the four-day school week.' *Education Finance and Policy* 10(3), 314–349

Andrietti, Vincenzo (2015) 'The causal effects of increased learning intensity on student achievement: Evidence from a natural experiment.' UC3M WP Economic Series 15-06, Universidad Carlos III de Madrid, June

_ (2016) 'The causal effects of an intensified curriculum on cognitive skills: Evidence from a natural experiment.' UC3M WP Economic Series 16-06, Universidad Carlos III de Madrid, April

Andrietti, Vincenzo, and Xuejuan Su (2016) 'Education curriculum and student achievement: Theory and evidence.' UC3M WP Economic Series 16-07, Universidad Carlos III de Madrid, April

Ashenfelter, Orley, and Alan Kruger (1994) 'Estimates of the economic return to schooling from a new sample of twins.' *American Economic Review* 84(5), 1157–1173

Aucejo, Esteban M., and Teresa Foy Romano (2016) 'Assessing the effect of school days and absences on test score performance.' *Economics of Education Review* 55, 70–87

Baumert, J. (2009) *Programme for International Student Assessment 2000 (PISA 2000). Version: 1* (IQB - Institut zur Qualitätsentwicklung im Bildungswesen. Datensatz. http://doi.org/10.5159/IQB_PISA_2000_v1)

Bellei, Cristian (2009) 'Does lengthening the school day increase students' academic achievement? results from a natural experiment in chile.' *Economics of Education Review* 28, 629–640

Bertrand, Marianne, Esther Duflo, and Sendhil Mullainathan (2004) 'How much should we trust differences-in-differences estimates?' *Quarterly Journal of Economics* 119(1), 249–275

Cameron, Colin A., and Douglas L. Miller (2015) 'A practitioner's guide to cluster-robust inference.' *Journal of Human Resources* 50(2), 317–372

Cameron, Colin A., Jonah G. Gelbach, and Douglas L. Miller (2008) 'Bootstrap-based improvements for inference with clustered errors.' *Review of Economics and Statistics* 90(3), 414–427

Card, David (2001) 'Estimating the return to schooling: Progress on some persistent econometric problems.' *Econometrica* 69(5), 1127–1160

Carlsson, Magnus, Gordon B. Dahl, Björn Öckert, and Dan-Olof Rooth (2015) 'The effect of schooling on cognitive skills.' *Review of Economics and Statistics* 97(3), 533–547

Clotfelter, Charles T., Helen F. Ladd, and Jacob L. Vigdor (2015) 'The aftermath of accelerating algebra: evidence from district policy initiatives.' *Journal of Human Resources* 50(1), 159–188

Cortes, Kalena E., Joshua S. Goodman, and Takako Nomi (2015) 'Intensive math instruction and educational attainment: Long-run impacts of double-dose algebra.' *Journal of Human Resources* 50(1), 108–158

Cuhna, Flavio, and James J. Heckman (2007) 'The technology of skill formation.' *American Economic Review* 97(2), 31–47

Dahmann, Sarah (2017) 'How does education improve cognitive skills? instructional time versus timing of instruction.' *Labour Economics* 47, 35–47. EALE conference issue 2016

Dahmann, Sarah, and Silke Anger (2014) 'The impact of education on personality. evidence from a german high school reform.' SOEP papers on Multidisciplinary Panel Data Research 658, DIW Berlin, May

Deming, David (2009) 'Early childhood intervention and life-cycle skill development: Evidence from head start.' *American Economic Journal: Applied Economics* 1(3), 111–134

Dougherty, Shaun M. (2015) 'Bridging the discontinuity in adolescent literacy? mixed evidence from a middle grades intervention.' *Education Finance and Policy* 10(2), 157–192

Fitzpatrick, Maria D., David Grissmer, and Sarah Hastedt (2011) 'What a difference a day makes: Estimating daily learning gains during kindergarten and first grade using a natural experiment.' *Economics of Education Review* 30, 267–279

Heckman, James, and Edward Vytlacil (1998) 'Instrumental variables methods for the correlated random coefficient model: Estimating the average rate of return to schooling when the return is correlated with schooling.' *Journal of Human Resouruces* 33(4), 974–987

Heckman, James J., Lance J. Lochner, and Petra E. Todd (2006) 'Earnings functions, rates of return and treatment effects: The mincer equation and beyond.' *Handbook of the Economics of Education* 1, 307–458

Huebener, Mathias, and Jan Marcus (2017) 'Moving up a gear: the impact of compressing instructional time into fewer years of schooling.' *Economics of Education Review* 58, 1–14

Huebener, Mathias, Susanne Kuger, and Jan Marcus (2017) 'Increasing instruction hours and the widening gap in student performance.' *Labour Economics* 47, 15–34. EALE conference issue 2016

Klieme, E. (2013) *Programme for International Student Assessment 2009 (PISA 2009). Version: 1* (IQB - Institut zur Qualitätsentwicklung im Bildungswesen. Datensatz. http://doi.org/10.5159/IQB_PISA_2009_v1)

KMK (1997-2014b) *Wochenpflichtstunden der Schülerinnen und Schüler - Statistiks 1997 bis 2014*

___ (2002-2012) *Allgemein bildende Schulen in Ganztagsform in den Ländern in der Bundesrepublik Deutschland - Statistik 2002 bis 2012*

___ (2014a) 'Schüler, klassen, lehrer und absolventin der schulen 2003 bis 2012.' Statistiche Veröffentlichungen der Kulturministerkonferenz. Dokumentation 204, March

Kraft, Matthew A. (2015) 'How to make additional time matter: Integrating individualized tutorials into an extended day.' *Education Finance and Policy* 10(1), 81–116

Krashinsky, Harry (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

Lavy, Victor (2015) 'Do differences in school's instruction time explain international achievement gaps in maths, science and language? evidence from developed and developing countries.' *Economic Journal* 125(588), F397–F424

Marcotte, Dave E. (2007) 'Schooling and test scores: A mother natural experiment.' *Economics of Education Review* 26, 629–640

Marcotte, Dave E., and Steven W. Hemelt (2008) 'Unscheduled school closing and student performance.' *Education Finance and Policy* 3(3), 316–338

Meyer, Tobias, and Stephan L. Thomsen (2016) 'How important is secondary school duration for post-school education decisions? evidence from a natural experiment.' *Journal of Human Capital* 10(1), 67–108

Morin, Louis-Philippe (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

OECD (2011) 'Germany: Once weak international standing prompts strong nationwide reforms for rapid improvemente.' In 'Strong performers and successful reformers in education: Lessons from PISA for the United States' (OECD Publishing) pp. 201–220

___ (2012) *PISA 2009 Technical Report* (OECD Publishing)

Parinduri, Rasyad A. (2014) 'Do children spend too much time in schools? evidence from a longer school year in indonesia.' *Economics of Education Review* 41, 89–104

Prenzel, M. (2007) *Programme for International Student Assessment 2003 (PISA 2003). Version: 1* (IQB - Institut zur Qualitätsentwicklung im Bildungswesen. Datensatz. http://doi.org/10.5159/IQB_PISA_2003_v1)

___ (2010) *Programme for International Student Assessment 2006 (PISA 2006). Version: 1* (IQB - Institut zur Qualitätentwicklung im Bildungswesen. Datensatz. http://doi.org/10.5159/IQB_PISA_2006_v1)

— (2015) *Programme for International Student Assessment 2012 (PISA 2012). Version: 1* (IQB - Institut zur Qualitätsentwicklung im Bildungswesen. Datensatz. http://doi.org/10.5159/IQB_PISA_2012_v1)

Rivkin, Steven G., and Jeffrey C. Schiman (2015) 'Instruction time, classroom quality, and academic achievement.' *Economic Journal* 125(588), F425–F448

Sims, David P. (2008) 'Strategic responses to school accountability measures: It's all in the timing.' *Economics of Education Review* 27, 58–68

Spinath, Birgit, Christine Eckert, and Ricarda Steinmayr (2014) 'Gender differences in school success: what are the roles of students' intelligence, personality, and motivation.' *Educational Research* 56(22), 230–243

Statistisches Bundesamt (2015) 'Bildungsausgaben: Ausgaben je schüler/-in 2012.' Technical Report, Statistisches Bundesamt, March

Taylor, Eric (2014) 'Spending more of the school day in math class: Evidence from a regression discontinuity in middle school.' *Journal of Public Economics* 117, 162–181

Thile, Hendrik, Stephan L. Thomsen, and Bettina Büttner (2014) 'Variation of learning intensity in late adolescence and the effect on personality traits.' *Journal of the Royal Statistical Society: Series A (Statistics in Society)* 177(4), 861–892

Fig. 1. PISA cohorts treatment status

PISA 2000-2003 cohorts

PISA 2006 cohorts

PISA 2009 cohorts

PISA 2012 cohorts
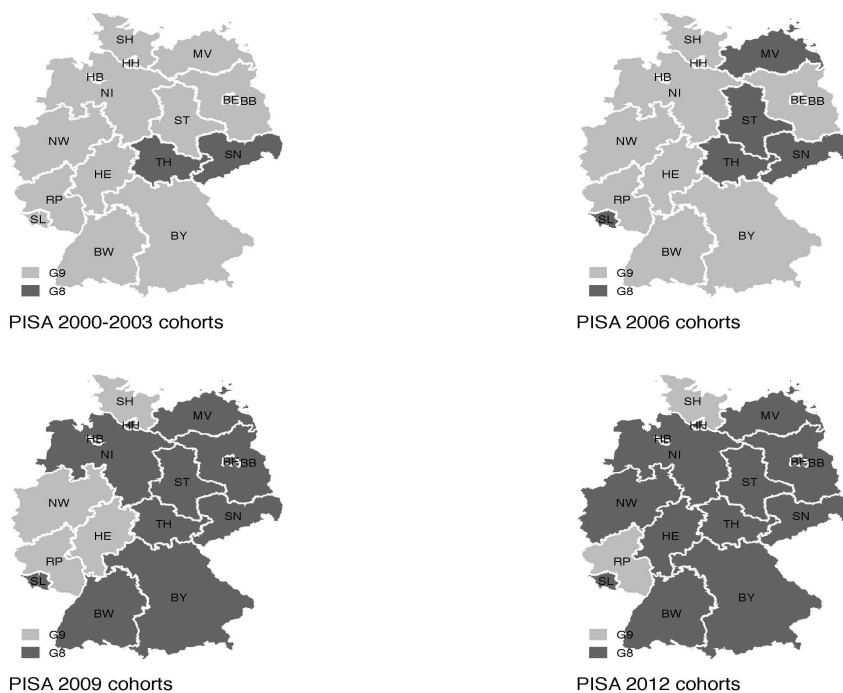
Legenda

G9: Control state, **G8**: Treatment state

BW: Baden-Württemberg, **BY**: Bavaria, **BE**: Berlin, **BB**: Brandenburg, **HB**: Bremen, **HH**: Hamburg, **HE**: Hesse, **MV**: Mecklenburg-Vorpommern, **NI**: Lower Saxony, **NW**: North Rhine-Westfalia, **RP**: Rhineland-Palatinate, **SL**: Saarland, **ST**: Saxony-Anhalt, **SN**: Saxony, **SH**: Schleswig-Holstein, **TH**: Thuringia.

Table 1. Summary statistics

Variable	Mean	SD
PISA scores		
Reading	572.13	55.51
Mathematics	578.39	58.26
Science	587.05	61.10
Student controls		
Female	0.53	0.50
Age (in months)	185.22	5.54
Parental education: Tertiary (ISCED ≥ 5)	0.62	0.49
Parental ISEI	59.25	17.34
Books in house: ¿100	0.58	0.49
Only child	0.29	0.45
Foreign born child	0.04	0.20
Foreign born parents	0.13	0.34
Foreign language spoken at home	0.04	0.20
School controls		
School enrollment	793.93	352.15
% of girls enrolled	49.42	15.07
Urban school	0.26	0.44
Private school	0.08	0.26
Student-teacher ratio	14.66	5.93
Student-computer ratio	26.78	62.84
Fraction of certified teachers	0.74	0.40
Fraction of part time teachers	0.35	0.19
Shortage of language arts teachers	0.06	0.24
Shortage of math teachers	0.20	0.40
Shortage of science teachers	0.24	0.43
Shortage of materials for instruction	0.23	0.42
Shortage of lab equipment	0.37	0.48
Shortage of library resources	0.31	0.46
Policy variables		
G8 reform	0.41	0.49
Years of treatment	1.61	2.30
Avg. weekly instruction hours (KMK: grades 5-9)	30.93	1.49

Notes: Descriptive statistics are weighted by PISA student sampling weights. The sample includes grade-9 academic-track students in PISA 2000–2012 with valid reading scores ($N = 33,996$). Statistics for mathematics and science scores are based on students with non-missing scores in mathematics ($N = 29,929$) and in science ($N = 30,202$), respectively.

Table 2. The G8 reform effects on weekly instruction hours

| | Official KMK timetables | | | | | | PISA grade 9 self-reported hours | |
	Avg. grades 5-9 (1)	Grade 5 (2)	Grade 6 (3)	Grade 7 (4)	Grade 8 (5)	Grade 9 (6)	Academic track (7)	Lower tracks (8)
G8 reform	2.473***	1.737***	1.235***	3.006***	3.137***	3.253***	2.291***	0.331
	(0.163)	(0.361)	(0.317)	(0.435)	(0.346)	(0.360)	(0.690)	(0.447)
Pre-reform avg. hours	29.66	28.33	28.84	30.14	30.28	30.73	29.53	26.74
Post-reform change (%)	8.4	6.1	4.3	10.0	10.4	10.6	12.1	0.01
Observations			33,996				14,025	28,402

Notes: Dependent variable in column 1: avg. year-week hours of instruction in academic-track grades 5-9. Dependent variables in columns 2 to 6: academic-track grade specific year-week hours of instruction. Dependent variables in columns 7 and 8: PISA grade nine year-week hours of instruction. All regression models are estimated using the baseline specification of equation (1) and PISA student weights. Standard errors clustered on state are reported in parentheses. ***, ** and * indicate significance at 1, 5 and 10 percent levels, respectively. The samples for columns 1-6 include academic-track ninth-graders in PISA 2000–2012 with a valid assessment in reading. The sample for column 7 includes academic-track ninth-graders in PISA 2003, 2009 and 2012 with non-missing values on the dependent variable. The sample for column 8 includes basic- and middle-track ninth-graders in PISA 2003, 2009 and 2012 with non-missing values on the dependent variable.

Table 3. The average effects of the G8 reform

	(1)	(2)	(3)
Panel A: Reading $(N = 33,996)$			
G8 reform	0.073***	0.081***	0.078***
	(0.022)	(0.021)	(0.022)
Years of treatment	0.013**	0.015***	0.014**
	(0.005)	(0.005)	(0.006)
Weekly instruction hours	0.031***	0.034***	0.034***
	(0.010)	(0.008)	(0.009)
Panel B: Math $(N = 29,929)$			
G8 reform	0.075*	0.081**	0.067**
	(0.044)	(0.035)	(0.032)
Years of treatment	0.015	0.016*	0.013*
	(0.010)	(0.008)	(0.008)
Weekly instruction hours	0.023*	0.024**	0.022**
	(0.014)	(0.011)	(0.010)
Panel C: Science $(N = 30,202)$			
G8 reform	0.088***	0.089***	0.085***
	(0.026)	(0.020)	(0.018)
Years of treatment	0.017***	0.016***	0.015***
	(0.006)	(0.005)	(0.004)
Weekly instruction hours	0.026**	0.027***	0.025***
	(0.011)	(0.009)	(0.008)
State and year fixed effects	✓	✓	✓
Student controls		✓	✓
School controls			✓

Notes: Regression models are estimated using different specifications of equation (1) and PISA student weights. Standard errors clustered at the state level are reported in parentheses. ∗∗∗, ∗∗ and ∗ indicate significance at 1, 5 and 10 percent levels respectively. The samples for panel A, B, and C include academic-track ninth-graders in PISA 2000–2012 with valid reading, math, and science scores, respectively.

Fig. 2. Inter-temporal effects of the G8 reform
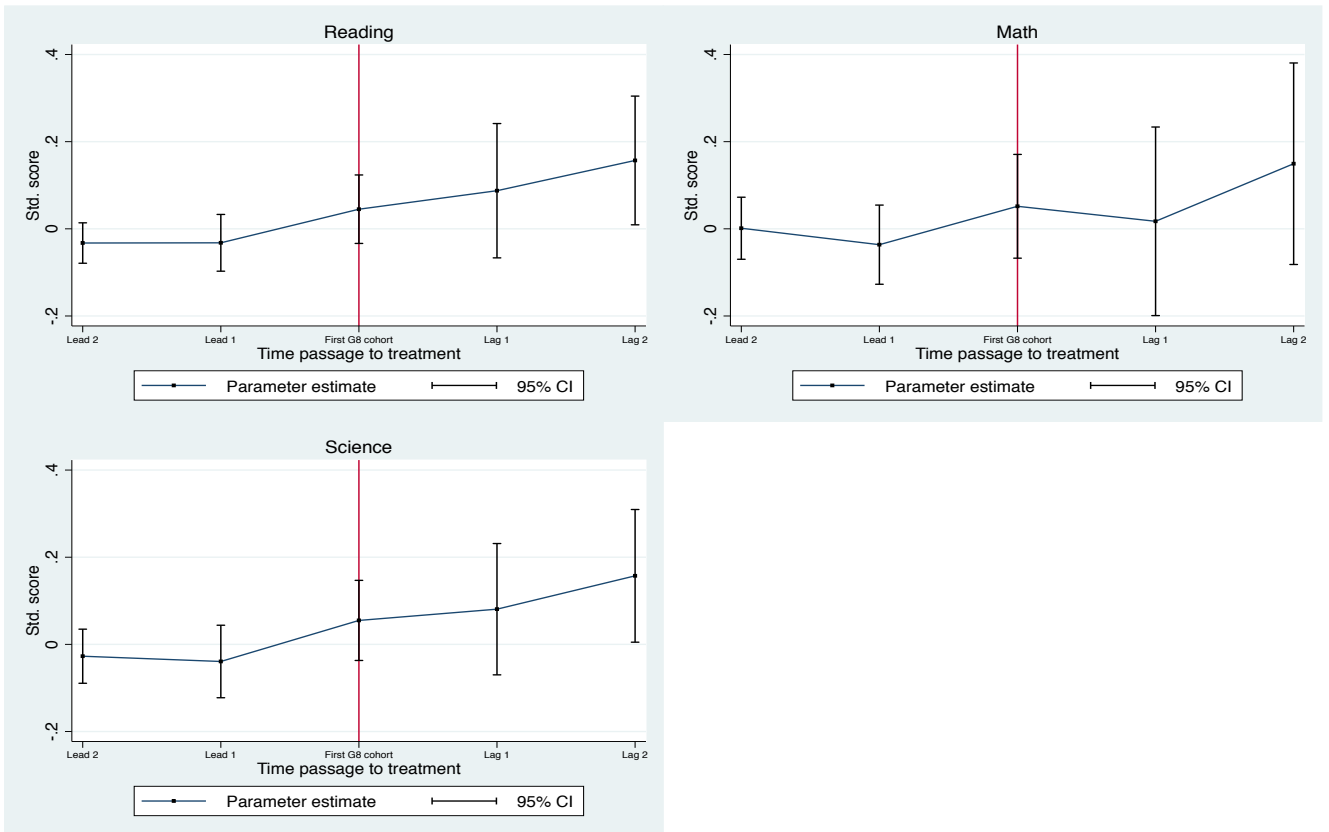
Source: Computations on PISA 2000-2012 pooled data (baseline specification, PISA student weights used)

Table 4. Sensitivity tests

	Main spec. (1)	DD lead (2)	Placebos lower-tracks (3)	DDD model (4)	State trends (5)	Switch to CEE (6)	All day schooling (7)	Double cohorts (8)
Panel A: Reading								
G8 reform	0.078***	-0.010	-0.017	0.115**	0.075**	0.069**	0.077***	0.081***
	(0.022)	(0.026)	(0.045)	(0.058)	(0.029)	(0.031)	(0.022)	(0.023)
Observations	33,996	33,996	57,748	72,053	33,996	33,996	33,996	33,996
Panel B: Math								
G8 reform	0.067**	-0.036	-0.022	0.092*	0.100***	0.061*	0.065**	0.065**
	(0.032)	(0.041)	(0.058)	(0.054)	(0.034)	(0.050)	(0.032)	(0.033)
Observations	29,929	29,929	50,542	63,289	29,929	29,929	29,929	29,929
Panel C: Science								
G8 reform	0.085***	-0.020	-0.001	0.094**	0.097***	0.084***	0.084***	0.089***
	(0.018)	(0.038)	(0.058)	(0.039)	(0.032)	(0.033)	(0.018)	(0.017)
Observations	30,202	30,202	50,988	63,886	30,202	30,202	30,202	30,202

Notes: All regression models are estimated using the main specification of equation (1) and PISA student weights. Standard errors clustered at the state level are reported in parentheses. $***$, $**$ and $*$ indicate significance at 1, 5 and 10 percent level respectively. The samples for columns 1, 2, and 5–8 include grade-9 academic-track students in PISA 2000–2012 with valid reading, math, and science scores respectively. The samples for column 3 include grade-9 basic- and middle-track students in PISA 2000–2012. The samples for column 4 include grade-9 middle- and academic-track students in PISA 2000–2012.

Table 5. Heterogeneous reform effects based on observed heterogeneity

	Reading (1)	Mathematics (2)	Science (3)
Panel A: Reform effects by gender			
G8 × Boy	0.003	0.081*	0.065***
	(0.024)	(0.043)	(0.020)
G8 × Girl	0.143***	0.056*	0.101***
	(0.025)	(0.030)	(0.020)
$P-value$ of t-test of difference	**0.00**	0.50	**0.05**
Observations	33,922	29,885	30,128
Panel B: Reform effects by age			
G8 × Age ≤ 1st tercile	0.059**	0.044	0.062**
	(0.027)	(0.034)	(0.026)
G8 × Age > 1st tercile	0.094***	0.087***	0.104***
	(0.027)	(0.033)	(0.021)
$P-value$ of t-test of difference	0.22	**0.02**	0.16
Observations	33,996	29,929	30,202
Panel C: Reform effects by parental immigration status			
G8 × Foreign born parents	0.019	0.007	-0.045
	(0.063)	(0.055)	(0.052)
G8 × German born parents	0.075***	0.060**	0.087***
	(0.020)	(0.030)	(0.018)
$P-value$ of t-test of difference	0.29	0.19	**0.00**
Observations	33,126	29,058	29,331
Panel D: Reform effects by parental education			
G8 × < Tertiary	0.099**	0.071	0.095***
	(0.038)	(0.044)	(0.033)
G8 × Tertiary	0.061***	0.052*	0.071***
	(0.023)	(0.027)	(0.023)
$P-value$ of t-test of difference	0.36	0.60	0.54
Observations	32,861	28,793	29,066
Panel E: Reform effects by parental ISEI			
G8 × ISEI ≤ 1st tercile	0.076**	0.040	0.073***
	(0.030)	(0.037)	(0.024)
G8 × ISEI > 1st tercile	0.079***	0.083**	0.091***
	(0.022)	(0.032)	(0.019)
$P-value$ of t-test of difference	0.88	**0.10**	0.40
Observations	33,680	29,612	29,885
Panel F: Reform effects by book at home			
G8 × Books at home: ≤ 100	0.050*	0.028	0.020
	(0.028)	(0.042)	(0.032)
G8 × Books at home: > 100	0.088***	0.076**	0.105***
	(0.022)	(0.031)	(0.019)
$P-value$ of t-test of difference	0.09	0.14	**0.01**
Observations	32,774	28,765	29,032

Notes: All estimated models are estimated using the main specification of equation (2) and PISA student weights. Standard errors clustered on state are reported in parentheses. ∗∗∗, ∗∗ and ∗ indicate significance at 1, 5 and 10 percent levels, respectively. When significant, p-values of the t-test of difference are reported in bold. The samples include ninth-graders in academic track schools from the pooled PISA 2000–2012 dataset with a valid assessment in reading, math, and science, and non-missing values in the characteristic considered.

Table 6. Effects on out-of-school study time/attendance and teachers/classroom quality

	Out-of-school study time (1)	Indexes of: Out-of-school class attendance (2)	Teacher quality (3)	Classroom quality (4)
G8 reform	-0.053 (0.157)	-0.052 (0.149)	0.056 (0.044)	0.201 (0.138)
Observations	7,973	9,162	29,081	21,574

Notes: All regression models are estimated using the main specification of equation (1) and PISA student weights. The samples in columns 1 and 2 include grade-9 academic-track students in PISA 2003 and 2012 with non-missing values on the variables used to build the dependent variable. Standard errors clustered on state are reported in parentheses. ∗∗∗, ∗∗ and ∗ indicate significance at 1, 5 and 10 percent levels, respectively. The sample in column 3 includes grade-9 academic-track students in PISA 2000 - 2012 with non-missing values on the variables used to build the dependent variable. The sample in column 4 includes grade-9 academic-track students in PISA 2000, 2003, 2009, and 2012 with non-missing values on the variables used to build the dependent variable.
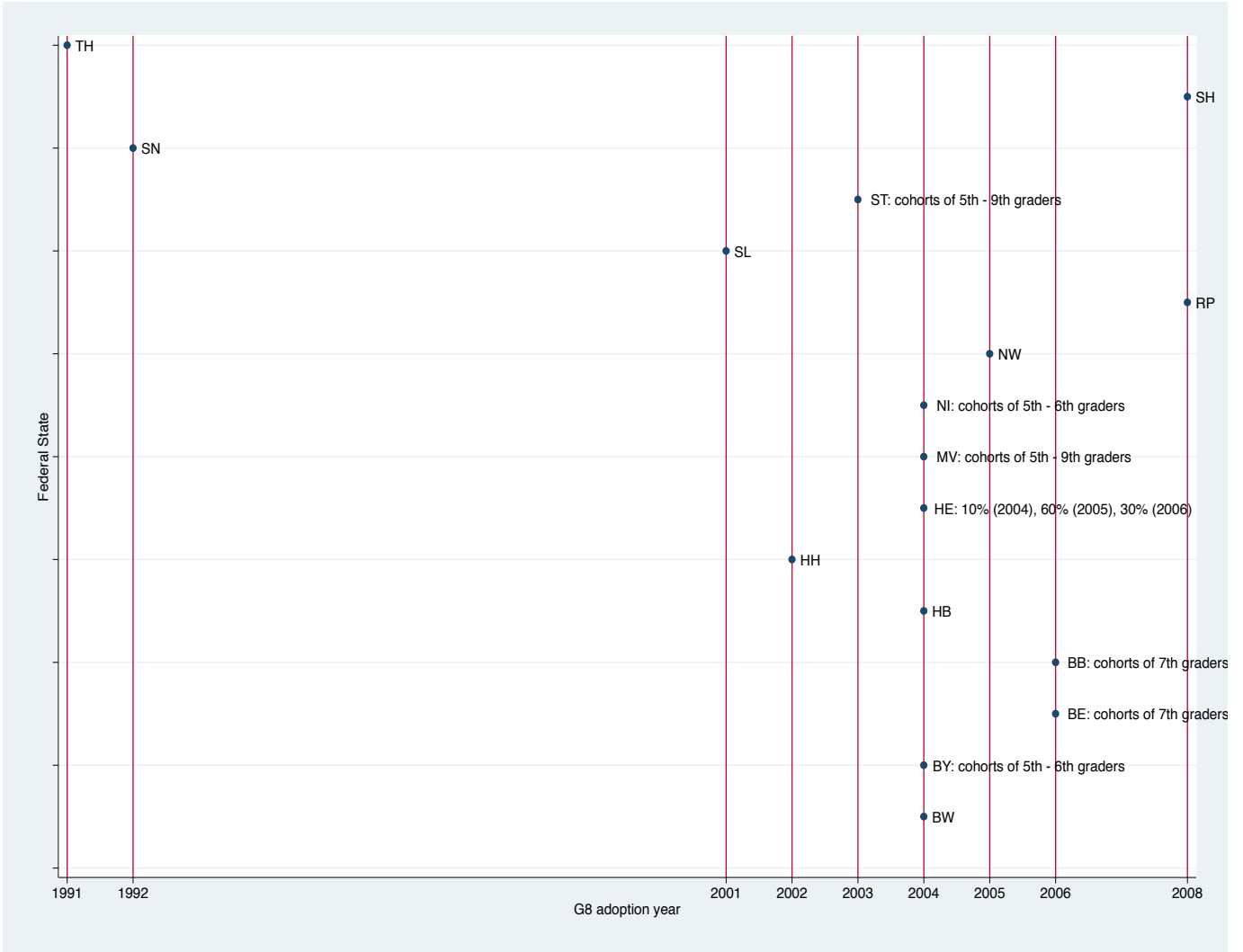
Table 7. Heterogeneous reform effects based on unobserved heterogeneity

	Quantiles								
	0.10	0.20	0.30	0.40	0.50	0.60	0.70	0.80	0.90
Panel A: Reading ($N = 33,996$)									
G8	0.038	0.054*	0.065**	0.075**	0.092***	0.098***	0.097***	0.102***	0.104***
	(0.036)	(0.032)	(0.029)	(0.031)	(0.025)	(0.023)	(0.025)	(0.032)	(0.040)
Panel B: Math ($N = 29,929$)									
G8	0.037	0.025	0.033	0.062***	0.072***	0.091***	0.093***	0.099***	0.084**
	(0.038)	(0.037)	(0.038)	(0.024)	(0.026)	(0.033)	(0.028)	(0.029)	(0.042)
Panel C: Science ($N = 30,202$)									
G8	0.059	0.071**	0.063**	0.077***	0.083***	0.096***	0.098***	0.102***	0.101**
	(0.047)	(0.036)	(0.028)	(0.025)	(0.028)	(0.024)	(0.024)	(0.029)	(0.048)

Notes: All regression models are estimated using the main specification of equation (3) and use PISA student weights. Conventional standard errors are reported in parentheses. ∗∗∗, ∗∗ and ∗ indicate significance at 1, 5 and 10 percent level respectively. The samples for Panel A, B, and C include grade-9 academic-track students in PISA 2000–2012 with valid reading, math, and science scores, respectively.

Appendix

Fig. A.1. G8 reform implementation timing and first treated cohorts

Table A.1. The G8 reform effects on student characteristics

	Gender (female) (1)	Age (months) (2)	Parental education: Tertiary (3)	Parental ISEI (4)	Books in house (¿100) (5)	Only child (6)	Parents born abroad (7)	Children born abroad (8)	Foreign language (9)
G8 reform	-0.009 (0.022)	0.203 (0.389)	-0.022 (0.019)	0.078 (0.988)	0.008 (0.026)	0.026 (0.018)	-0.028 (0.031)	-0.011 (0.008)	-0.004 (0.008)
Observations	33,922	33,996	33,100	33,680	32,774	32,979	33,126	32,794	31,602

Notes: Dependent variables in columns 1–9: student controls included in the main specification of equation (1). All regression models are estimated using the baseline specification of equation (1) and PISA student weights. Standard errors clustered at the state level are reported in parentheses. $***$, $**$ and $*$ indicate significance at 1, 5 and 10 percent level respectively. The samples include grade-9 academic-track students in PISA 2000–2012 with valid reading scores and non-missing values of the dependent variable.

Table A.2. The G8 reform effects on school characteristics

	School enrollment (1)	% girls enrolled (2)	Urban school (3)	Private school (4)	Student-teacher ratio (5)	Student-computer ratio (6)	Fraction teachers certified (7)	Fraction teachers part time (8)	Shortage of teachers German (9)	math (10)	science (11)	Shortage of resources textbooks (12)	labs (13)	library (14)
G8 reform	39.948 (49.033)	-0.617 (1.591)	-0.037 (0.072)	-0.009 (0.024)	-1.144 (1.255)	3.201 (6.389)	0.088 (0.069)	-0.044 (0.052)	0.071 (0.049)	0.019 (0.135)	0.186 (0.144)	-0.027 (0.113)	0.006 (0.086)	0.150 (0.095)
Observations	32,234	32,175	32,955	32,956	31,319	30,821	28,986	31,060	32,758	32,897	32,901	32,845	32,837	32,772

Notes: Dependent variables in columns 1–14: school controls included in the main specification of equation (1). All regression models are estimated using the baseline specification of equation (1) and PISA student weights. Standard errors clustered at the state level are reported in parentheses. ∗ ∗ ∗, ∗∗ and ∗ indicate significance at 1, 5 and 10 percent level respectively. The samples include grade-9 academic-track students in PISA 2000–2012 with valid reading scores and non-missing values on the dependent variable.

Department of Economics, University of Alberta
Working Paper Series

2017-03: The Voting Rights of Ex-Felons and Election Outcomes in the United States – **Klumpp, T.,** Mialon, H., Williams, M.

2017-02: Does the Design of a Fiscal Rule Matter for Welfare? – **Landon, S., Smith, C.**

2017-01: Carbon Pricing with an Output Subsidy under Imperfect Competition: The Case of Alberta's Restructured Electricity Market – **Brown, D., Eckert, A., Eckert, H.**

2016-18: Monetary Policy Tradeoffs Between Financial Stability and Price Stability – **Shukayev, M.,** Ueberfeldt, A.

2016-17: Managing Risk Taking with Interest Rate Policy and Macroprudential Regulations – Cociuba, S., **Shukayev, M.,** Ueberfeldt, A.

2016-16: On the Role of Maximum Demand Charges in the Presence of Distributed Generation Resources – **Brown, D.,** Sappington, D.

2016-15: Implementing Cross-Border Interbank Lending in BoC-GEM-FIN – **Shukayev, M.,** Toktamyssov, A.

2016-14: The Effects of Early Pregnancy on Education, Physical Health and Mental Distress: Evidence from Mexico – **Gunes, P.,** Tsaneva, M.

2016-13: An Equilibrium Selection Theory of Monopolization – **Eckert, A., Klumpp, T., Su, X.**

2016-12: Education Curriculum and Student Achievement: Theory and Evidence – Andrietti, V., **Su, X.**

2016-11: Poverty and Aging – **Marchand, J.,** Smeeding, T.

2016-10: Local Labor Markets and Natural Resources: A Synthesis of the Literature – **Marchand, J.,** Weber, J.

2016-09: Accounting for Firm Exit and Loss of Variety in the Welfare Cost of Regulations – **Andersen, D.**

2016-08: Analyzing the Impact of Electricity Market Structure Changes and Mergers: The Importance of Forward Commitments – **Brown, D., Eckert, A.**

2016-07: Credibility of History-Dependent Monetary Policies and Macroeconomic Instability – Cateau, G., **Shukayev, M.**

2016-06: Electricity Market Mergers with Endogenous Forward Contracting – **Brown, D., Eckert, A.**

2016-05: Thinking about Minimum Wage Increases in Alberta: Theoretically, Empirically, and Regionally – **Marchand, J.**

2016-04: Economic and Socio-Demographic Determinants of Child Nutritional Status in Egypt: A Comprehensive Analysis using Quantile Regression Approach– **Sharaf,M.,** Rashad,A.

2016-03: Regional Inequalities in Child Malnutrition in Egypt, Jordan, and Yemen: A Blinder-Oaxaca Decomposition Analysis – Rashad, A., **Sharaf, M.**

2016-02: Collateralized Borrowing and Risk Taking at Low Interest Rates – Cociuba, S., **Shukayev, M.,** Ueberfeldt, A.

2016-01: Optimal Policies to Promote Efficient Distributed Generation of Electricity – **Brown, D.,** Sappington, D.

2015-18: Departure and Promotion of U.S. Patent Examiners: Do Patent Characteristics Matter? - **Langinier, C.,** Lluis, S.

2015-17: Socioeconomic Inequalities in Infant Mortality in Egypt: Analyzing Trends between 1995 and 2014 – **Sharaf, M.,** Rashad, A.

2015-16: Does Economic Growth Reduce Child Malnutrition in Egypt? New Evidence from National Demographic and Health Survey – Rashad, A., **Sharaf, M.**

2015-15: The Labor Market and School Finance Effects of the Texas Shale Boom on Teacher Quality and Student Achievement – **Marchand, J.,** Weber, J.