How Does Education Improve Cognitive Skills?

Instructional Time versus Timing of Instruction

Sarah C. Dahmann (DIW Berlin)*

March 7, 2016

Abstract

This paper investigates two mechanisms through which education may affect cognitive skills in adolescence, exploiting a school reform carried out at the state level in Germany as a quasi-natural experiment to identify causal effects: Between 2001 and 2007, years at academic-track high school were reduced by one, leaving the overall curriculum unchanged. First, I exploit the variation over time and across states to identify the effect of an increase in class hours on same-aged students' intelligence scores, using data on seventeen year-olds from the German Socio-Economic Panel. Second, I investigate the influence of earlier instruction at younger ages, using data from the German National Educational Panel Study on high school graduates' competences. The results suggest that overall, secondary education impacts students' crystallized cognitive skills in adolescence especially through instructional time rather than through age-distinct timing of instruction. However, they also reveal that increasing instructional time aggravates gender differences in numeracy.

Keywords: Cognitive Skills, Crystallized Intelligence, Fluid Intelligence, Skill Formation, Education, High School Reform
JEL Classification: I21, I28, J24

*I thank Silke Anger, Stijn Baert, Friederike von Haaren, Susanne Kuger, Henning Lohmann, Bettina Siflinger, C. Katharina Spieß, Stefan C. Wolter, and seminar participants at DIW Berlin, the International Young Scholar SOEP Symposium, the Spring Meeting of Young Economists, the Essen Health Conference, the International Workshop of Applied Economics of Education, the Annual Conference of the European Economic Association, and the Annual Conference of the Verein für Socialpolitik for helpful comments and discussions. Furthermore, I gratefully acknowledge funding from the German National Academic Foundation. Correspondence to: Sarah Dahmann, German Institute for Economic Research (DIW Berlin), Mohrenstr. 58, 10117 Berlin, Germany, Tel: +49-30-89789-461, E-mail: sdahmann@diw.de.

1 Introduction

Cognitive skills are important determinants of many economic and social outcomes. At a macro level, cognitive skills in a population are strongly related to a country's economic growth (Hanushek and Woessmann, 2008). At a micro level, higher cognitive skills are associated with, among others, increased health and better old-age functioning mental abilities, and they are also linked to higher wages (see e.g. Heckman et al., 2006, or Heineck and Anger, 2010) and better education. The latter association is, however, a two-way relationship. On the one hand, individuals with higher cognitive abilities are likely to be better educated as they choose more often to continue education or easier meet access requirements to enter, e.g., university after secondary education. On the other hand, education itself also improves cognitive skills. Most studies use changes in compulsory schooling laws as an exogenous variation to identify causal positive effects of an additional year of schooling on cognition (e.g. Banks and Mazzonna, 2012). However, they do not provide evidence on the underlying mechanisms. This paper, therefore, investigates the roles of instructional time and timing of instruction as two potentially important channels through which secondary education may affect cognitive skills.

My research question is two-fold, which will be addressed in two analyses: First, I assess the impact of an increase in instructional time – dedicated to corresponding additional curriculum – on cognitive skills of adolescents in Germany. Second, I investigate whether the timing of instruction influences cognitive skill development, i.e. whether the allocation of class hours at a younger age changes cognitive skills, keeping the level of education constant. In both analyses, I allow for gender heterogeneity in the effects to further investigate whether such educational changes are mitigating or aggravating factors for gender skill differences (see e.g. Wigfield et al., 2002, for a review on gender skill differences).

To address these research questions, I exploit a reform in German high schools implemented between 2001 and 2007 that shortened total years of schooling from thirteen to twelve, leaving the overall curriculum unchanged. As a result, the number of weekly class hours significantly increased. Hence, while still in school, affected students have covered a greater share of the overall curriculum than non-affected students of the same age. I use this intensified curriculum as an exogenous increase in the instructional quantity received up to the age of seventeen and exploit the variation over time and region in the implementation of the reform to identify its causal effect on adolescents' cognitive skills. Using rich data on adolescents from the German Socio-Economic Panel (SOEP) study, difference-in-differences estimates show that the reform improved crystallized intelligence, while fluid intelligence remained largely unaffected. However, the former impact significantly differs by gender: whereas male students' scores improved especially in numerical skills, female students' skills hardly improved at all. These results indicate the importance of instructional time as a mechanism in education improving cognitive skills, but also reveal its aggravating role in gender skill differences. I further use the variation in the age at which students received instruction as a quasi natural-experiment to investigate the impact of educational timing on students' competences. Using extensive data from the German National Educational Panel Study (NEPS) for the federal state of Baden-Wuerttemberg on students in their final grade, estimations suggest that the earlier knowledge transfer did not significantly alter the development of competences among students affected by the reform. Here, the potential benefit of early investment and age effects seem to offset each other. As a result, students affected by the reform catch up with their non-affected counterparts in terms of their competences by the time of graduation, apart from potential age effects resulting in slightly decreased fluid intelligence scores.

Next, I describe the theoretical background and existing literature. When explaining the high school reform in more detail, I elaborate on potential channels and anticipated effects. After a description of data and empirical strategy, I present the results. I test the robustness of the findings in several sensitivity analyses, before I conclude discussing the implications.

2 Theoretical Basis and Previous Literature

Theoretical Basis. Cognitive skills shape a variety of later-life outcomes. Together with non-cognitive skills, they form an important part of an individual's human capital as they constitute personal skills. A common approach to describe the formation and development of such skills is proposed by Cunha and Heckman (2007). They argue that an individual's present stock of skills depends on his or her past stock of skills, previous investment, and environmental factors. More specifically, they suggest the following model:

$$\theta_{t+1} = f_t(\theta_t, I_t, h) \tag{1}$$

where a vector of skill stocks at age $t+1$, θ_{t+1}, depends in some positive functional form $f(\cdot)$ on the past vector of skills (with initial endowment θ_1), on the investment in period t, I_t, and on parental, or more generally environmental, characteristics, h. In this model, Cunha and Heckman propose a multiplier effect driven by two mechanisms, *self-productivity* and *dynamic complementarity*. Self-productivity occurs whenever $\partial f_t(\theta_t, I_t, h)/\partial\theta_t > 0$. This implies that skills persist such that higher skills at one point in time create higher skills in the subsequent period, and is not restricted to one and the same skill but also includes cross effects between different skills. Dynamic complementarity occurs whenever $\partial^2 f_t(\theta_t, I_t, h)/\partial\theta_t\partial I_t' > 0$ implying that the productivity of investment is increasing with higher existing skills. Cunha and Heckman (2008) test and verify both propositions empirically. Hence, the resulting multiplier effect suggests that investments are most productive in early stages in life, making childhood the critical period for skill formation. Skills may, therefore, be malleable through e.g. educational interventions, especially at an early stage in life. However, there are important differences across dimensions of skills to distinguish.

Cognitive skills are usually distinguished into different facets. Two major ones in the empirical literature are fluid intelligence and crystallized intelligence.[1] Fluid intelligence

[1] For a more detailed overview, see for example Baltes (1993) who describes fluid abilities as the *fluid-like mechanics of intelligence* and crystallized abilities as the *crystallized pragmatics of intelligence*.

relates to innate abilities that people are genetically endowed with. These include, for example, the ability to reason, the level of comprehension, or the capability of processing information, and are usually not influenced to a great extent by environmental factors. Crystallized intelligence, in contrast, denotes explicitly or implicitly learned knowledge or behavior. Therefore, it covers any specific knowledge of facts, for example, as well as learned behavioral traits such as the ability to read or calculate. Unlike fluid intelligence, crystallized intelligence is determined through environmental factors like education or upbringing. Several studies show that education indeed improves the crystallized component of cognitive skills, both in the short- and long-run.

Previous Literature. For Scandinavian countries, several studies use data on males between the ages 18 and 20 from military cognitive assessment tests to identify short-term effects: Brinch and Galloway (2012) use an increase in compulsory schooling from seven to nine years in Norway between 1955 and 1972. Their difference-in-differences estimates and their instrumental variable results suggest positive returns and translate an additional year of schooling into an increase of 3.7 IQ points[2]. For Sweden, Carlsson et al. (2015) exploit a random variation in test dates to find that one additional year of schooling leads to an increase in crystallized intelligence of up to 0.21 standard deviations. Fluid intelligence does not seem to be affected by schooling, but rather positively by age. Instrumenting schooling and initial IQ, Falch and Massih (2011) find cognitive returns to one additional year of schooling between 2.9 and 3.8 IQ points for the Swedish population in Malmö that enrolled in the military in 1947 and 1948. Cascio and Lewis (2006) use data from the 1979 National Longitudinal Survey of Youth (NLSY79) to estimate returns to schooling on the Armed Forces Qualifying Test (AFQT) scores of males and females aged 15 to 19 years in the United States. Exploiting variation in the date of birth and school entry regulations, they find positive effects of 0.32 standard deviations; however only for racial and ethnic minorities. Setting up a regression discontinuity design to analyze the long-term effects of a compulsory schooling reform in England, Banks and Mazzonna (2012) find an increase in memory functioning of between 0.35 and 0.6 standard deviations

[2]Note that generally IQ scales are defined to have mean 100 and standard deviation 15. An increase of 3.7 IQ points thus roughly corresponds to a 0.25 standard deviation increase.

among males and females older than 50. In turn, executive functioning only increased for males, with effect sizes ranging from 0.37 to 0.63 of a standard deviation. Using SHARE data on Austria, Czech Republic, Denmark, France, Germany, and Italy, Schneeweis et al. (2014) exploit the variation in compulsory schooling across the different countries to investigate cognitive ability of individuals older than 50. They find positive effects of 0.1 standard deviations of one additional year of schooling on memory functioning as well as some evidence on the reduction in cognitive decline in terms of verbal fluency through schooling. Furthermore, their effect sizes are generally larger in magnitude for males. Lastly, Kamhöfer and Schmitz (2015) investigate the long-term impact of education in Germany on word fluency among males and females born between 1940 to 1970 using data from the German Socio-Economic Panel (SOEP) study in 2006. They use different instruments for schooling to estimate local average treatment effects but find no effects. However, their outcome is limited in the sense that it is a single-edged view on cognition as it does not cover further dimensions of cognition next to word fluency, and that it is based on an ultra-short test that is conducted in only 90 seconds. Furthermore, weak instruments may be a threat to their identification, while age and cohort specific effects cannot be disentangled, which may confound their results.

Hence, with the exception of the study on Germany by Kamhöfer and Schmitz (2015), all studies clearly find substantial positive effects of an additional year of schooling on cognitive abilities. To establish effect causality, most of these analyses exploit a change in overall school duration by one year. Still, the underlying mechanisms remain unresolved. However, for policy conclusions, it is critical to understand whether there are driving forces beyond overall school duration behind this relationship. While school duration, *per se*, cannot be changed infinitely, the existence of underlying channels would open new possibilities for decision makers to target cognitive ability when designing educational policies.

A change in school duration may have different consequences related to skill formation. On the one hand, an additional year of schooling may induce a larger curriculum to cover, i.e. constitute a direct increase in time and material of instruction. On the other hand, a change in the overall years of schooling may as well only lead to a redistribution of covered material and instruction over the different grades, i.e. over different age spans of the students. While the former constitutes a direct increase in investment I_t in equation (1), the latter implies a shift in the timing of investment I_t. Both may, therefore, impact cognitive skills: on the one hand, keeping age and past skills constant, an increase in investment, i.e. an increase in instructional quantity, may directly improve cognitive abilities. On the other hand, keeping overall instruction quantity constant, the age at which instruction for a given topic is received may influence cognitive abilities as well. Here controversial mechanisms could interact, where earlier instruction is assumed to increase returns from later investments according to Cunha and Heckman (2007) and thereby improve cognitive skills, and because skills are more malleable at younger ages, but later instruction could benefit from maturity or time required to digest instruction.[3]

It therefore still remains to investigate whether either instructional time or timing of instruction drive the positive relationship between schooling and cognitive skills or whether it is both. To the best of my knowledge, this study is the first to investigate and disentangle these two mechanisms. To identify causal effects, I use a unique variation in the German schooling system that allows me to conduct two separate analyses to provide a complete picture. First, keeping age constant, the causal effect of an increase in instructional time is identified. Second, keeping the educational level constant, the role of instructional timing and age is analyzed. In addition, this study extends the literature on Germany, especially given the puzzle that Kamhöfer and Schmitz (2015) find no effects while for all other countries investigated there exist positive cognitive returns to education. Furthermore, the rich datasets contain extensive tests of cognitive ability allowing for different cognitive dimensions to be distinguished. In addition, the inclusion of female respondents enables the investigation of gender heterogeneity to uncover whether education may be a mitigating or aggravating factor for gender skill differences. In developmental and educational psychology gender differences in abilities have been a long-standing focus: Wigfield et al. (2002) summarize in their review, that originally

[3]Existing literature related to any of these particular mechanisms will be elaborated on in Section 3.2 when discussing anticipated effects of the reform.

girls exhibited higher verbal skills and performance, while boys showed higher mathematical and spatial abilities. Although they note that these gender differences declined over time, gaps in mathematical and physical abilities favoring boys persist. Investigating a sample of high school seniors performing the ACT Assessment Mathematics Usage Test, Doolittle (1989) finds that particularly in geometry and reasoning items, females perform worse than males. These gender differences in subject-specific dimensions of abilities are of particular interest given the ongoing policy efforts to promote female participation in STEM – Science, Technology, Engineering and Mathematics – subjects (see e.g. OECD, 2014). To some extent, these differences may be attributable to biological factors (see e.g. Lynn, 1994), but also to environmental aspects like gender stereotypes influencing self-perception of abilities (Jacobs et al., 2002) and education (Ellison and Swanson, 2012). Understanding the influence of schooling in these gender skill differences is, therefore, crucial to develop adequate educational policies.

3 The German High School Reform

3.1 Institutional Background and Change

In Germany, educational policy is at the responsibility of the federal states. In all cases, however, children enter elementary school at the age of six and continue on to secondary education usually after four years.[4] Secondary education in Germany is provided at three different levels, listed in ascending order by their level of education provided: *Hauptschule* (basic track), *Realschule* (intermediate track) and *Gymnasium* (upper track). Of these three, only successful completion of *Gymnasium* (henceforth referred to as academic-track or simply high school) leads to the *Abitur*, the university entrance qualification. With a share of 34.4% of all German secondary students in the 2012/13 academic year attending *Gymnasium*, it is the most attended type of secondary school (Malecki et al., 2014).

Typically, high school lasted nine years, implying a total of thirteen years of schooling.

[4]Exceptions hereto are Berlin, Brandenburg and Mecklenburg-West Pomerania, where, in general, elementary school encompasses the first *six* grades. The assignment to different types of secondary school therefore takes place at grade seven.

Starting in 2001, several German states reduced this time at high school by one year, enabling graduation after completing only twelve years of schooling.[5,6] However, the overall curriculum remained unchanged.[7] As a result, weekly class hours significantly increased and school days prolonged. The increase of, on average, 3.7 class hours per week constitutes an increase of 12.5% of overall week hours.[8] The allocation of this increase in workload to different grades is determined on a state and school level, but grades seven to nine are usually most affected. Although the reform was implemented across almost the entire country, the timing of the introduction differs by state. An overview of the implementation of the reform by federal state is depicted in Figure A.1 in the Appendix.

3.2 Anticipated Effects of the Reform on Cognitive Skills

The reform may affect students' cognitive skills through several channels. I aim to disentangle two important mechanisms: the effect of an increase in instructional time (keeping age constant) and the role of earlier instruction (keeping reached educational level constant).[9] According to the Cunha and Heckman (2007) skill formation model, both cases should be assumed to lead to higher cognitive skills. Still, it is an empirical question whether and to what extent these mechanisms can be verified to lead to higher cogni-

[5] An extensive discussion on the reasons for this reform can be found in Dahmann and Anger (2014).

[6] A similar educational policy change took place in Ontario, Canada, in 1999: Krashinsky (2014) finds that students with one year less of high school perform significantly worse at university than their counterparts in terms of grades. Unlike the German high school reform, however, this change effectively reduced the curriculum taught as the number of years was reduced along with the number of courses available to students. The German setting is therefore unique in the sense that school duration was altered but the overall curriculum was not.

[7] From grade five through receiving the *Abitur*, 265 year-week hours must be completed. Year-week hours are the number of class hours in each year that are summed up over all years. This restriction was kept even while reducing high school duration from nine to eight years. In the states where elementary school encompasses the first six grades, the reform reduced time at high school from seven to six years. The year-weak hours requirement holds in the same way, however, as for other states, counting class hours from grade five onwards.

[8] With nine years at high school, the average week hours amounted to 265/9=29.44 hours; with only eight years at high school they increased to 265/8=33.13 hours.

[9] Note that an increase in instructional time when filled with additional content but keeping age constant, naturally leads to the introduction of certain parts of the curriculum at younger ages. The first mechanism is therefore not perfectly to disentangle from earlier timing. However, the results reveal that the earlier timing hardly yields effects. Further, I consider a substantial increase in instructional time by more than 800 class hours, which can therefore be expected to clearly dominate in the analysis on this mechanism.

tion, especially during adolescence. Even if they hold, further aspects may hinder or offset their positive effects. Therefore, I conduct two separate analyses; the results of which shed light on the mechanisms behind the relation between schooling and cognitive abilities.

In both cases, effect heterogeneity based on initial skill differences could be expected: as proposed by Cunha and Heckman (2007), returns to investment increase with higher existing skills. Hence, students outperforming others may benefit in particular from the increased instructional quantity. In this context, gender differences are of particular interest, as descriptives show that prior to the reform male students scored higher than females in most domains of cognitive abilities among the seventeen year-olds (SOEP Sample, see Table A.4) and the high school graduates (NEPS Sample, see Table A.7). With respect to these domains, I therefore hypothesize that male students benefit more from the increase in educational investment yielding larger positive reform effects for male than for female students. Further sources for gender-specific reform effects may be that male and female students are affected differently by school-related aspects which may change along with the reform, like peer pressure (Tinklin, 2003) and school excellence (Ellison and Swanson, 2012). Additionally, the returns to non-cognitive skills in terms of school achievement differ by gender (Spinath et al., 2010). Hence, changes in personality traits induced by the reform, as found by Dahmann and Anger (2014), may be beneficial for females' cognitive achievement while harmful for that of males, or vice versa. For these reasons, I allow for gender heterogeneity in the reform's effect in both analyses.

The first analysis compares same-aged, i.e. seventeen-year-old students, where the students affected by the reform have accumulated significantly more class hours, which were filled with corresponding additional curriculum. This increase in instructional quantity should especially raise crystallized measures of intelligence, while fluid intelligence is generally assumed unaffected.[10] Few studies similarly investigate the impact of class

[10]Note that it is not possible to completely separate these two dimensions of intelligence in a test environment. As soon as e.g. speed is introduced to give specific knowledge, fluid and crystallized skills are required simultaneously. Furthermore, Baltes (1993, p. 581) notes that in practice, crystallized and fluid skills interact and that, in addition "the pragmatics [crystallized intelligence] always build on the mechanics [fluid intelligence]". Further, Cunha and Heckman (2007) assume that skills are cross-fertilizing, i.e. that changes in one domain of skills foster changes in another domain.

hours, but on more curriculum-oriented achievement tests: Andrietti (2015) estimates the same reform's effects on PISA test scores of ninth-graders. Therefore, these students have been affected by the reform for less time compared to my sample of investigation of seventeen year-olds. Still, he finds positive effects in reading, mathematical, and science literacy skills, with the first being driven by female students. Machin and McNally (2008) employ difference-in-differences estimation to evaluate the introduction of a literacy hour in English elementary schools. They find that devoting one hour per day on English literacy along with changing the structure and content of teaching increases students' rank in reading and English skills by 2 to 3 percentage points. Taylor (2014) uses a fuzzy regression discontinuity design to investigate the effect of increasing the share of class hours spent in math classes in sixth grade at Miami-Dade County Public Schools. He finds that math achievement rises by 0.16 to 0.18 standard deviations, but that effects fade with time passed since the remediation course. Contrary to the German High School Reform, these latter policy changes do not constitute an increase in *overall* instructional time. Keeping the length of school days constant, increases in instructional time in one subject may therefore come at the cost of other subjects. Cortes et al. (2015) analyze a policy change in Chicago Public Schools that doubled the amount of time devoted to

Table 1: Anticipated Effects of Increased Instruction Quantity on Cognitive Skills of Same-aged Students

Cognitive Skills	Effect	Potential Channel
Crystallized	+	Increase in instructional time with corresponding additional taught curriculum until age 17
	(+	Multiplier effect of earlier instruction)
	0	Not malleable anymore between ages 10 and 17
	0	Additional knowledge taught cannot be absorbed anymore
	0	Increase in formal instructional time substitutes informal learning
	−	Long school days come at cost of extra-curricular activities which may be important for cognitive skill development (direct effects, or indirect effects through changes in non-cognitive skills)
Fluid	0	Not malleable
	+/−	Indirect effects through changes in crystallized intelligence

algebra for low skilled ninth-graders. Using a regression discontinuity design, they find positive effects on achievement test scores and further outcomes. Using heavy snowfall as an exogenous variation in the number of school days that students in Maryland could attend, Marcotte (2007) finds that students with less instructional time performed significantly worse on the Maryland School Performance Assessment Program (MSAP) exams. Different to the German high school reform, these two studies investigate the effect of an increase in instructional time keeping the curriculum constant. The increased (or decreased) time therefore serves for more (or less) repetition and practice of the same content, i.e. decelerates (or accelerates) the speed of learning during each class hour. In contrast, the reform analyzed in this study provides a unique setting in which an increase in instructional time implies both an increase in class hours along with the corresponding increase in the curriculum taught. In this case, it may be that either the additional knowledge taught cannot be absorbed by the students[11] or simply that cognitive skills are no longer malleable at this age in adolescence, bringing no particular change in cognition at all. Lastly, the increase in formal instructional time may substitute informal cognitive stimulating activities or come at the cost of further, e.g. non-cognitive, skills or extracurricular activities important for skill development, offsetting the positive effects on cognition or even negatively impacting them.[12] An overview of these anticipated effects can be found in Table 1.

The second analysis compares students in their final year of high school, although at different biological ages. At this point in time both students affected by the reform and students not affected have reached the same educational level, accumulating the same number of class hours. However, students affected by the reform have received this instruction at a relatively younger age. According to Cunha and Heckman (2007) this earlier investment – presumably leading to higher cognitive skills at an earlier stage in life, which will be tested in the first analysis – increases a person's stock of skills at an earlier stage making any investment thereafter even more productive. As a result, students

[11] Whether this is the case may especially differ between distinct types of students, as e.g. students with lower initial skills may have more difficulties with keeping up at the new pace.

[12] Dahmann and Anger (2014) show that the reform indeed had an effect on some personality traits. The participation in extracurricular activities seems however not to be affected (see Table 5).

Table 2: Anticipated Effects of Earlier Instruction on Cognitive Skills of Students in Final Grade

Cognitive Skills	Effect	Potential Channel
Crystallized	+	Multiplier effect making instruction more productive (if indeed the increase in instructional time increased cognitive skills at a younger age)
	0	Formal instruction substitutes informal learning thereby not changing the timing of learning
	–	Biological age effects
	–	Required maturity
Fluid	0	Not malleable
	–	Biological age effects
	+/–	Indirect effects through changes in crystallized intelligence

affected by the reform may have acquired higher cognitive skills through this multiplier effect of early investment, at least in crystallized dimensions of cognition. Again, if the instruction only substitutes already present learning, thereby not altering the timing of learning, there should be no effect. However, while attending their final year of high school, these students are one year younger than those students not affected by the reform, which may have negative consequences for cognition, including both crystallized and fluid dimensions.[13] Furthermore, students may lack the maturity to digest particular subjects at a younger age making instruction less productive, as found by Clotfelter et al. (2015). Exploiting a policy shift in some school districts of North Carolina, they find that accelerating the introduction of algebra coursework into eighth grade has significant negative impacts on students' performances in algebra and the follow-up geometry course. Furthermore, they find that low performing students are harmed the most, further increasing inequality. Unlike this policy change, the German high school reform is not bound to any particular subject, rather applying to the complete high school curriculum. Table 2 summarizes the hypothesized effects.

[13]See, for example, Baltes (1987), who illustrates the life-span development of cognitive abilities: Both, crystallized and fluid intelligence, peak close to the age of 25; however, crystallized ability remains relatively stable thereafter, whereas fluid ability decreases with age. Important to note is though, that up to the early 20-years, both domains of intelligence increase with age, mostly irrespective of the environment.

4 Data

To investigate both potential mechanisms, I conduct two analyses. The first compares same-aged students to evaluate the impact of an increase in instructional time on cognitive skills: Students affected by the reform have accumulated a higher number of class hours than students prior to the reform. The second investigates a sample of students at the same educational level, i.e. at completion of secondary school, to identify the role of earlier knowledge transfer implied by the reform for affected students. By nature, these two samples differ and pose different requirements to the underlying dataset to enable identification. Therefore, the two analyses are based on different datasets.

4.1 The German Socio-Economic Panel (SOEP)

The first analysis is based on data from the German Socio-Economic Panel (SOEP) study, which is a representative household panel surveyed annually (Wagner et al., 2007) with information on around 30,000 individuals in almost 15,000 households in the 2013 wave.[14] In addition to various individual and household characteristics, including family background and childhood environment, the SOEP includes cognitive potential measures for different subsamples since 2006. The cognitive abilities of adolescents aged seventeen[15], who respond to the SOEP youth questionnaire, are assessed in every wave starting in 2006. Thus, I use the 2006 through 2013 waves, including all adolescent respondents aged seventeen[16] who attend *Gymnasium* in my sample. To identify whether a student is affected by the reform, I use the information on the federal state of residence and the year of school entry. In case information on the latter is not provided, the year of school entry is imputed from the date of birth. As Saxony and Thuringia established a

[14]This paper uses data from the Socio-Economic Panel (SOEP), data for years 1984-2013, version 30, SOEP, 2014, `doi:10.5684/soep.v30`.

[15]In the SOEP, adolescents are interviewed in the year they *turn* seventeen. Thus, the age of seventeen results from defining age simply as the difference between the year of survey and the year of birth. Note, however, that their *real* age at the time of the interview is either sixteen or seventeen, depending on their date of birth and the date of the interview.

[16]In 2006, when the test of cognitive abilities was conducted for the first time, adolescent respondents from the 2004 and 2005 waves were also tested. To increase the sample size, I also include these individuals (aged eighteen and nineteen) in my preferred specification. Birth year dummies control for potential age effects. Still, a robustness check including only seventeen year-olds is conducted to confirm the results.

twelve-year-school system before Germany's reunification, I consider all students in these two states as affected.[17] I exclude students from Rhineland-Palatinate where the reform has not been implemented state-wide. To avoid adding noise to the amount and level of education received by the subjects, I exclude all students who have repeated any grade.[18] Lastly, I restrict the sample to those who successfully completed the cognitive assessment test and have valid information on their background and family characteristics. The final sample consists of 723 students, of whom 288 are affected by the reform.

Cognitive skill measures. In the SOEP adolescent questionnaire, cognitive skills are measured through a short form of the I-S-T 2000 R (see Amthauer et al., 2001) that takes 30 minutes. This test consists of three parts, each with 20 questions.[19] The first part consists of word analogies and measures verbal skills: participants are asked to find a matching word according to a specific rule. In the second part numerical skills are measured, where the respondent has to fill in the correct arithmetic operators in incomplete equations. Together, these two (verbal and numerical) tasks record crystallized intelligence as they reflect an individual's explicitly learned competences. In contrast, the third task serves to measure fluid intelligence: here three abstract figures are displayed according to a specific rule with participants asked to pick a fourth figure from five proposed figures. On each of these three test components adolescents answer as many questions as possible, in the given amount of time. The scores then measure the number of correct answers (out of 20 possible questions). To facilitate the interpretation of results, I standardize all scores separately by gender to mean zero and variance one. Summary statistics are given in Tables A.3 and A.4 and the development of cognitive skills over time is graphically illustrated in Figure A.2 in the Appendix.

Other variables. To account for individual characteristics that may also influence

[17] In a robustness check, students from these two states are excluded.

[18] This procedure would threaten my identification if repetition rates changed with the reform. However, Huebener and Marcus (2015) find that repetition rates up to grade nine remained unchanged by this reform. Even though they changed in the final years before graduation, for my sample of seventeen-year olds repetition rates should therefore be rather similar before and after the reform. In my sample only 55 students drop due to grade repetition; of which 29 are affected by the reform and 26 are not. Still, I include these grade repeaters in a robustness check to confirm the results.

[19] For an extensive overview of the measurement and assessment of adolescents' cognitive potential in the SOEP see Schupp and Herrmann (2009) and Richter et al. (2013).

cognitive abilities, I control for several pre-reform characteristics in my preferred specification. These include socio-economic and demographic variables like gender, migration background, and when they were born[20]. Furthermore I capture a student's previous performance by the teacher's recommendation after elementary school.[21] Family variables include parental characteristics based on education, work status and occupational status, and also capture whether a student grew up with only one parent.[22,23]

4.2 The German National Educational Panel Study (NEPS)

The second analysis is based on data from the German National Educational Panel Study (NEPS), which is a longitudinal dataset aimed at mapping competence development and learning environment over the life cycle. It follows a multicohort sequence design starting with more than 60,000 target persons from six cohorts (Blossfeld et al., 2011). In addition to these six original cohorts, it includes a cross-sectional additional study in the German federal state of Baden-Wuerttemberg, which targeted students at academic-track high school in their final year.[24] Baden-Wuerttemberg is the third largest federal state in Germany, both in terms of area and population, with a share of 34% of students at

[20]Specifically, this measures whether students were born in the first or in the second half of the year and, thereby, controls for the grade in school they attend at the date of the interview: As the vast majority of the interviews (77% of my sample) are administered during the first quarter of the year in which adolescents turn seventeen, students born between July and December usually attend grade ten. In contrast, students born between January and June usually enter school comparatively young and are, therefore, on average, one grade more advanced at the time of the interview.

[21]At the end of elementary school, teachers recommend one of the different secondary school tracks to the students' parents based on their perception of the student's performance and potential. I classify students as low-performing if they received a recommendation for either *Realschule* or *Hauptschule*, i.e. the intermediate and lower secondary tracks. Even though this recommendation is not equally binding across all federal states, the number of these students attending *Gymnasium* nevertheless, is naturally very low.

[22]For a subsample also information on the number of siblings and birth order, which both may be relevant for cognitive skills, is available. However, to avoid a loss in sample size, I do not include these variables in the main specification. Estimations controlling for being the oldest child and being an only child are not reported but confirm the findings. As well, controlling for household size, which is often found to also correlate with cognition, confirms the findings. As this information is however problematic since it is not measured before students are affected by the reform, I do not include it in the specification.

[23]A detailed overview of all variables included in the analysis is depicted in Table A.1.

[24]This paper uses data from the National Educational Panel Study (NEPS): Additional Study Baden-Wuerttemberg, `doi:10.5157/NEPS:BW:3.0.0`. From 2008 to 2013, NEPS data were collected as part of the Framework Programme for the Promotion of Empirical Educational Research funded by the German Federal Ministry of Education and Research (BMBF). As of 2014, the NEPS survey is carried out by the Leibniz Institute for Educational Trajectories (LIfBi) at the University of Bamberg in cooperation with a nationwide network.

Gymnasium in 2012/2013 (Malecki et al., 2014) which corresponds almost perfectly to the German average of 34.4%. In Baden-Wuerttemberg, the last cohort not affected by the high school reform and the first affected cohort both graduated in 2012. Therefore, in 2012, the NEPS target population consisted of this double sized graduation class. Hence, I use the 2012 wave, including all respondents who attend the final grade of *Gymnasium* in my sample. Again, I exclude all students who repeated a grade. The final sample consists of 2,128 students, of whom 1,113 are affected by the reform.

Cognitive skill measures. Cognitive abilities are measured through an extensive 2 hours 40 minutes test covering different educational dimensions. Reflecting explicitly learned knowledge, a 30 minutes achievement test in mathematics constitutes a measure of crystallized intelligence.[25] On this test, students are given a set of 21 questions. Most questions are multiple choice, with others partly answered in an open format. Therefore, a weighted maximum likelihood estimate (WLE; Warm, 1989) based on the test items constitutes an individual's measure of mathematical ability. Fluid intelligence is covered by measures of general cognitive abilities, i.e. perceptual speed and reasoning. Perceptual speed is assessed by a picture symbol test where respondents are required to enter correct figures for the preset symbols according to an answer key (see Lang et al., 2007), with a total of 3×31 items to be solved in 3×30 seconds. Reasoning is measured in the same way as figural skills in the SOEP adolescent questionnaire: Based on Raven's matrices, students fill in a missing geometrical element that fits the other elements of the matrix, in a total of 3×4 cases with 3×3 minutes time. For both of these fluid cognitive skills measures, the total score is calculated as the sum of correctly solved items. Again, for both crystallized and fluid measures of cognitive ability, I standardize all scores separately by gender to have mean zero and variance one.

Other variables. In addition to achievement tests, the survey also includes further individual and school characteristics. Whether or not a student is affected by the reform is given in the survey. Individual characteristics include demographics as gender

[25]Furthermore, physics, biology and English are assessed in this NEPS study. However, ability scores in these subjects based on the respective achievement tests, as of November 2015, are not yet released. Once released, the data will provide an interesting extension to this work that acknowledges the multi-dimensionality of crystallized intelligence.

and migration background. Furthermore, the number of books at home, parental education, the father's work classification and the mother's occupational status characterize a student's socio-economic background and home environment. In addition, a survey conducted at the school level allows me to further control for school characteristics in some specifications, including school size, the share of students and teachers with a migration background, as well as stress factors caused by the implementation of the reform. Stress factors are areas such as resources and organization where the headmaster reported to have had particular difficulty when implementing the reform.[26]

5 Empirical Strategy

I exploit the German high school reform introduced in almost all federal states between 2001 and 2007 as a quasi-natural experiment to identify a causal effect of education on cognition. The control group consists of students who entered high school before the reform was introduced and, therefore, graduate after nine years of high school. In contrast, the treatment group consists of students entering high school after the implementation of the reform and thus graduating after only eight years.

5.1 Estimation using SOEP

For the first analysis, all students in the selected SOEP sample usually attend either grade ten or grade eleven at the time of the interview, thus having spent about either $5^1/_2$ or $6^1/_2$ years in high school at the time of the interview.[27] However, the amount of education received during this time differs between the control group and the treatment group, as the reform provides an exogenous variation in the number of class hours attended at the time of the interview. Students affected by the reform should have accumulated at least between 800 and 945 class hours of education more, on average, than their non-affected

[26] A detailed overview of all variables included in the analysis is depicted in Table A.2.

[27] Recall that SOEP adolescents usually attend either grade ten (in case they were born between July and December) or grade eleven (in case they were born between January and June) at the time of the interview in the year they turn seventeen.

counterparts at the same age.[28]

In a difference-in-differences framework, I assess the impact of this intensified curriculum on cognitive skills. I exploit the variation over time and region to identify causal effects estimating the following equation:

$$y_{ist,17} = \alpha \text{REFORM}_{st} + X_i\beta + \sum_s \gamma_s \text{STATE}_s + \sum_t \delta_t \text{YEAR}_t + \varepsilon_{ist}, \quad (2)$$

where $y_{ist,17}$ is a measure of cognitive ability at age 17 of student i living in state s and born in year t. The variable of interest, $REFORM_{st}$ indicates whether students belong to the treatment or the control group. It equals 1 if students entering school in state s in year $t+6$ (or $t+7$ respectively, depending on their month of birth)[29] are affected by the reform when entering high school (i.e. belong to the treatment group) and 0 otherwise (i.e. belong to the control group). $STATE_s$ is a set of state dummies and $YEAR_t$ dummies indicate the year of birth. X_i is a vector of pre-reform individual characteristics, including the student's own demographic characteristics as well as childhood and family variables. The error terms, ε_{ist}, are clustered at the state level.[30]

Crucial to the identification of the prime parameter of interest, α, as a causal impact of education on cognitive skills, is the assumption that in absence of the reform, cognitive skills of students from the treatment group and of students from the control group do not differ significantly, i.e. $\alpha = 0$. This implies that cognitive skills develop similarly among students across states. While this so-called Common Trend Assumption is not testable, it should be reasonable and not too restrictive in this case: here, students of the same school type are compared across different states. Since students likely select into different school types based on initial abilities and socio-economic background, I assume

[28] The numbers are calculated as follows: (265/8-265/9)[average weekly increase in class hours due to the reform]*39.5[weeks of school per year]*5.5[years in high school so far] (or *6.5 years respectively).

[29] For the sample under consideration, the cutoff date is equal among all federal states: June 30. Hence, students born between January and June entered first grade six years after their year of birth, and students born between July and December entered first grade seven years after their year of birth.

[30] To account for the small number of clusters, it may be necessary to use wild cluster bootstrapped standard errors (see Cameron et al., 2008). The estimations, however, show that the wild cluster bootstrap leads to even slightly lower standard errors if different at all. I therefore report the usual standard errors without bootstrapping, as it is the more conservative estimation method in this case. Bootstrapped estimation results can be found in Table A.9.

that students differ severely across different types of secondary schools and, therefore, also in their development of competences. In contrast, students *at high school*, but living in different states, can be expected to possess similar initial characteristics. Still, I allow for state-specific linear time trends in a robustness check.

Furthermore, self-selection should not be possible, thus enabling a causal interpretation of the results. As the reform was introduced state-wide at the same time,[31] students did not have a choice on whether to be affected by the reform or not. It could however be, that students attended high school in a different state that had not yet introduced the reform. This however imposes high moving costs for the entire family and, therefore, seems unlikely.[32] Hence, selection *within* the sample, i.e. between treatment and control group, was hardly possible.[33] Still, the robustness checks in section 7 will consider subsamples where this possibility is ruled out to verify the estimation results. In contrast, selection *out of* the sample is possible by attending a different type of secondary school instead of high school. With the exception of comprehensive schools in some states, the graduation from the lower and intermediate secondary school does not lead to the *Abitur*.[34] Given the ever growing importance of educational certificates on the labor market, this is a far-reaching decision and can, therefore, be assumed to be relatively rarely a direct implication of the newly introduced reform. Indeed, Huebener and Marcus (2015) find that high school entry and graduation rates are not affected at all by the reform. Still, a robustness check deals with the existence of comprehensive schools in a state, and a Placebo estimation on students from the other school types suggests no such out-of-sample selection.

[31]The only exceptions to this are Rhineland Palatinate, which is excluded from my analysis, and Hesse, where the reform was gradually introduced for students newly entering high school in the school years 2004/05 (10% of all schools), 2005/06 (60%), and 2006/07 (30%). Therefore, I only include students from Hesse who entered high school in 2003 or earlier and were, therefore, not affected, as well as students who entered high school in 2006 or later and were, therefore, affected.

[32]Indeed, 94.5% of the students in my sample still live in the same town since their childhood.

[33]As the reform has only been announced and implemented *after* these students had entered elementary school already, students could only change their grade level by repeating or skipping a class. Note that when being in the first cohort affected, skipping a class to escape the reform would be pointless as one would graduate in the exact same year as originally. The same holds true when repeating a class with the reverse aim. In any other cohort which was neither the first affected nor the last non-affected cohort, no such behavior would have changed the treatment status.

[34]The *Abitur* can still be obtained later through evening classes. However, the vast majority is obtained at *Gymnasium*.

Lastly, the timing of the implementation of the reform may be related to certain state specific characteristics. According to Black et al. (2005) it is not crucial for my identification for the reform to be unrelated to these as I control for state fixed effects in the analysis. Nevertheless, see Dahmann and Anger (2014) for an investigation into the reform's implementation. They find that it is unrelated to the percentage of high school students in a state's population, to whether the government is conservative, to whether the next state elections were scheduled for 2001/2002, or to the state's GDP per capita. There is suggestive evidence that states with a higher median age of residents implemented the reform slightly earlier; an artifact related to the older population in East German states.

5.2 Estimation using NEPS

For the second analysis, all students in the selected NEPS sample are in their final grade of high school. Students affected by the reform are therefore in grade twelve, while students not affected by the reform are attending grade thirteen. However, both groups are at the same educational stage as they have accumulated the same quantity of instruction received, i.e. the accumulated number of class hours. At each school these groups attend the classes together during their final year.[35] However, the students affected by the reform have received this educational instruction at a younger age compared to their non-affected counterparts. As this earlier instruction may have increased their cognitive skills at a younger age,[36] students affected by the reform may possess a higher stock of skills than same-aged students not affected by the reform. According to Cunha and Heckman (2007), this higher stock of already existing skills is assumed to make investment more productive. Hence the instruction received thereafter may have larger benefits. Due to this multiplier effect, students affected by the reform may possess higher cognitive skills than non-affected students at the end of secondary education. However, affected students

[35] Along with the reform of shortening high school, Baden-Wuerttemberg revised the curriculum to move from an input-oriented teaching to an output-oriented teaching focusing on achieving educational standards. In the two final years, however, both affected and non-affected students attend the same classes and are, therefore, subject to the exact same curriculum and type of teaching.

[36] This is the case if the reform induced higher cognitive skills for same-aged students. Looking at the age of seventeen, this can be deduced from the results described in section 6.1.

are also, on average, one year younger than non-affected students, at the end of secondary education. Hence it is an empirical question whether and to what extent the multiplier effect can be found or is offset by potential biological age effects. Furthermore, both mechanisms may affect crystallized and fluid intelligence differently.

To estimate this relationship between the timing of education and students' cognitive skills, I estimate a reduced version of equation (2) as there is no variation over time and across states in this sample:

$$y_{ij} = \alpha \text{REFORM}_i + X_i\beta + \varepsilon_{ij}, \tag{3}$$

where y_{ij} is a measure of cognitive ability of person i at school j, $REFORM_i$ is a dummy indicating whether person i is affected by the reform ($REFORM_i = 1$) or not ($REFORM_i = 0$), and X_i is a vector of individual characteristics. The error terms, ε_{ij}, are clustered at the school level.[37] The prime parameter of interest, α, indicates the role of the timing of the instruction received in students' cognitive skill development: a positive α could prove the existence of the multiplier effect, proposed by Cunha and Heckman (2007) for early life interventions, even in adolescence. A negative α in turn, could stem from potential age effects. Of course both effects may not be present or may offset each other, thus yielding inconclusive results.

To interpret this relationship as causal, it is crucial that the reform indeed constitutes a quasi-natural experiment. For this to hold, no selection should be possible while the treatment and control groups should be comparable in terms of both observable and unobservable characteristics. As the reform was introduced state-wide at the same time, students did not have a choice on whether to be affected by the reform or not; hence selection to treatment or control group *within* this sample can be ruled out.[38] However, selection *out of* the sample may have been possible, but is again assumed to be unlikely.[39]

[37]The number of different schools amounts to 48 which suffices for standard inference to be valid according to Cameron et al. (2008) who propose to use wild cluster bootstrap only in case of 30 clusters or less.

[38]Similar to the discussion on selection in section 5.1, again changing from the treatment to the control group (or vice versa) would only be possible by skipping a class (repeating a class). However, as both groups end up in the same graduating classes, this is pointless.

[39]This could happen if students drop out of this double cohort either by repeating or skipping a class.

Furthermore, Table A.8 shows that treatment and control group are comparable with respect to the selected observable characteristics. While the comparability of unobserved characteristics in turn cannot be tested formally, the well-balanced observed characteristics together with the absence of self-selection point to the validity of this assumption.

6 Results

6.1 The Impact of Instructional Time on Cognitive Skills

Estimation results of equation (2) are presented in Table 3.[40] The few salient effects of the individual control variables that reveal statistical significance[41] are in line with expectations: in particular students who received the recommendation *not* to follow onto high school after elementary school, possess significantly less skills throughout all domains. Furthermore, students with a migration background show comparably less fluid skills and students with a working-class family background lack behind in both verbal and figural skills.

It can be seen that, on average, the increased instruction quantity induced by the reform has no significant impact on students' cognitive abilities.[42] Yet, the positive sign across all dimensions is in line with theory. However, when allowing for effect heterogeneity by gender (see Table 4), it can be seen that these are driven by improvements among male students: While there is virtually no effect among female students, male students'

The latter is extremely rare and the former would not have changed the fact of being affected by the reform; hence it would be selection unrelated to the implementation of the reform, which should therefore not pose a threat to the identification. A further possibility would be to move to a different state where the reform had not yet been introduced. Involving high moving costs for the entire family, this option seems highly unlikely. So does choosing an alternative secondary school track instead of high school. Nonetheless see 5.1 for a more extensive discussion of potential selection.

[40]As the data on adolescents in the SOEP is a pooled cross-section over several waves, no appropriate weights exist. To account for possible over- or underrepresentation of certain demographic groups, I include dummies for each SOEP subsample in this, as well as all following, estimations instead. The different SOEP subsamples correspond to newly entering groups in the survey, partly with a demographic focus as target.

[41]Given the small sample size, note that the lack of statistical significance does not *prove* but instead *fails to reject* that there is no effect. Further, note that as measures of cognitive skills are standardized separately by gender, the coefficient of female mechanically should equal zero. This does however not imply the absence of gender differences. These differences (prior and post reform) are presented in Table A.4.

[42]Note that the absence of statistical significance does not necessarily imply a zero effect. Given the relatively small sample size, a lack of statistical power can naturally be expected.

Table 3: Average Effects of the Reform

	Outcome Variables: Cognitive Skills		
	Crystallized		Fluid
	Verbal	Numerical	Figural
Reform	0.070	0.143	0.091
	(0.095)	(0.135)	(0.098)
Female	0.006	0.012	0.007
	(0.055)	(0.079)	(0.053)
Migration background	-0.175	-0.205	-0.296***
	(0.171)	(0.163)	(0.060)
Born January–June	0.069	0.011	0.090
	(0.060)	(0.086)	(0.100)
Low-performing student	-0.414***	-0.382***	-0.264**
	(0.080)	(0.098)	(0.093)
Rural area	-0.004	0.018	0.045
	(0.098)	(0.063)	(0.073)
High parental education	-0.007	-0.098	-0.014
	(0.097)	(0.100)	(0.075)
Working-class father	-0.278***	0.081	-0.137*
	(0.083)	(0.098)	(0.074)
Working mother	0.038	0.082	0.014
	(0.071)	(0.099)	(0.080)
Single parent	0.017	0.056	0.100
	(0.105)	(0.072)	(0.119)
R^2	0.099	0.080	0.101
Observations	723	723	723

Notes: SOEPv30 waves 2006 to 2013. OLS regressions. Further, a maximum set of state dummies, year of birth dummies, dummies for the different SOEP subsamples, and a constant are included. Standard errors, reported in parentheses, are clustered at the state level. * p<0.1, ** p<0.05, *** p<0.01.

numerical skills clearly improve by more than a quarter of a standard deviation following the reform. Given that male students in this sample outperformed female students in numerical abilities even before the introduction of the reform,[43] it is notable that here the increased instructional quantity is an aggravating, rather than mitigating, factor for gender skill differences. The initial dominance of male students in numerical skills also

[43] A potential reason for this observed gender difference could be a greater variability among males compared to females, i.e. a larger share of male adolescents scoring particularly low and high on the skill assessment (see e.g. Hedges and Nowell, 1995). In this case, including only students at academic-track high school in the analysis, who presumably possess higher skills, mechanically raises skill averages among male compared to female students. In my case however, the inspection of students at *all* types of secondary school shows that the greater variability hypothesis is not supported by this data but rather that males outperform females in verbal and numerical skills across all types of secondary school.

Table 4: Heterogeneous Effects of the Reform by Gender

	Outcome Variables: Cognitive Skills		
	Crystallized		Fluid
	Verbal	Numerical	Figural
Reform	0.097	0.297**	0.144
	(0.107)	(0.126)	(0.112)
Reform*Female	-0.053	-0.296***	-0.103
	(0.157)	(0.095)	(0.103)
R^2	0.099	0.085	0.102
Observations	723	723	723

Notes: SOEPv30 waves 2006 to 2013. OLS regressions. Further, a maximum set of state dummies, year of birth dummies, dummies for the different SOEP subsamples, and a constant are included. Individual characteristics controlled for include female, migration background, born January–June, low-performing student, rural area, high parental education, working-class father, working mother, and single parent. Standard errors, reported in parentheses, are clustered at the state level. * p<0.1, ** p<0.05, *** p<0.01.

provides two potential reasons for why instructional time may be more beneficial for male students than for female students. On the one hand, the initially higher numerical skills can shape male preferences for choosing mathematically demanding subjects.[44] On the other hand, the high initial numerical skills especially among male students constitute a higher stock of already existing skills. According to the assumption of Cunha and Heckman (2007) that investment is more productive when existing skills are higher, the increase in instructional quantity especially benefits those with already higher skills, i.e. male students in the domain of numerical skills. Like this, education seems to improve especially domains of skills with comparative advantages among the respective group of students. With regard to the remaining domains, verbal as well as figural skills show slightly better improvements among male students than among females following the reform. For both groups, however, effects fail to exhibit statistical significance. One reason could be that numerical skills are acquired primarily in school, while verbal skills are also promoted outside school through leisure-time reading and social interactions. In this case,

[44] In Germany the choice of the major fields of study is only possible in the last two years of high school. After the reform, therefore, this choice takes place one year earlier than it did before the reform. However, as interviews are largely conducted in the first quarter of the year that a student turns seventeen years old, most students in my sample are not able to choose major fields yet (or did so only very recently). Even if they were, note that as mathematics belongs to the core subjects it cannot be eliminated by any of the students. The same holds true for German literature.

the increase in formal instruction could have replaced the informal acquisition of verbal skills; thereby, yielding no significant changes in this domain but rather in numerical skills.

Following similar lines of reasoning as for gender differences, further effect heterogeneities could enhance inequality. With a particular focus on disadvantaged students, defined by socio-economic or migration background, no such differences are found, however. Neither does prior performance significantly differentiate reform effects, nor are students born in the first half of the calender year affected differently than those born in the second half of the year, although they differ by construction in the grade in school they attend at the time of the interview.[45]

Potential Channels. So far, I find positive effects of the increase in instructional time on students' cognitive abilities with only increases in numerical skills being statistically significant. More specifically, numerical skills increased for males only, severely aggravating the gender skill gap in this domain, which was prevalent even prior to the reform. To further investigate potential driving factors in these gender-specific effects, I estimate the reform's impact on further outcomes to uncover whether male and female students reacted differently to the reform (see Table 5).

First, I analyze whether leisure-time activities, which may be related to cognitive skill development, were crowded out by the reform. The estimates show that there were no such effects neither on music, sports, reading, or technical work.[46] Hence, although the increase in instructional time came with longer school days, relevant after-school activities do not seem to have been crowded out. Important to note is that these results hold for both male and female students, such that leisure-time behavior cannot account for the gender heterogeneity in the reform's effects.

Second, I analyze whether outcomes related to additional investment in the students' performance at school changed following the reform. These may hint at whether students

[45] Estimation results are available from the author upon request.

[46] The outcome variable for music and sports in each case refers to participating in this activity *at all*. Investigating the frequency of the activity (at least daily or at least once a week) instead, does not alter the results. The outcome variable for reading and technical work or programming in each case refers to participating in this activity *at least once a week*. Here as well, investigating the reform's effects on daily participation does not show a different pattern.

Table 5: Channels – Effects of the Reform on Leisure-time Activities, Paid Tutor Lessons and Parental Involvement

	Outcome Variables: Participation in Activity			
	Music	Sport	Reading	Tech. work
Reform	0.081	-0.036	-0.033	0.002
	(0.115)	(0.052)	(0.092)	(0.113)
Reform*Female	0.033	-0.015	0.096	-0.058
	(0.085)	(0.053)	(0.085)	(0.078)
Observations	723	723	723	721
	Outcome Variables: Lessons and Parental Involvement			
	Paid tutor	Parents:		
	lessons	Interest	Homework	Problems
Reform	-0.100	-0.038	-0.003	-0.127
	(0.075)	(0.075)	(0.067)	(0.105)
Reform*Female	0.126**	0.057	-0.002	0.065
	(0.045)	(0.078)	(0.056)	(0.100)
Observations	722	722	721	722

Notes: SOEPv30 waves 2006 to 2013. OLS regressions. Further, a maximum set of state dummies, year of birth dummies, dummies for the different SOEP subsamples, and a constant are included. Individual characteristics controlled for include female, migration background, born January–June, low-performing student, rural area, high parental education, working-class father, working mother, and single parent. Standard errors, reported in parentheses, are clustered at the state level. * p<0.1, ** p<0.05, *** p<0.01.

cope with the accelerated learning through the increased instruction by the same age or rely on additional resources. In particular I look at the use of paid tutor lessons, and parental involvement in the student's educational outcomes. The latter is measured by general interest of parents in their child's school performance, their help with homework, as well as problems arising between children and parents as a result of disagreements over studies. While there is no impact of the reform on these measures of parental effort, there is a significant increase in the utilization of paid tutor lessons among female students. Compared to their male counterparts, the share of female students using tutors rises by 12.6 percentage points. These results indicate that female students may have problems in absorbing the additional curriculum. Therefore, they might not benefit from the increase in instructional time as male students do, which could explain the gender differences in the effects of the reform that were especially salient with respect to numerical skills.

6.2 The Impact of Timing of Instruction on Cognitive Skills

Estimation results of equation (3) are presented in Table 6.[47] The effects of the individual control variables on cognitive abilities are in line with expectations: Whereas children with a migration background score relatively lower on the competence measures, students from high socio-economic backgrounds, measured by parental education or the number of books at home, have acquired higher skills especially in the crystallized domain of competences. Further, students born in the first half of the year slightly outperform students born in the second half of the year in mathematics and perceptual speed. Entering elementary school at a relatively younger age, these students may have benefited from the earlier instruction from first grade onwards and a stimulating environment with more older classmates.

However, the reform exhibits no statistically significant effects on mathematical competence, as one dimension of crystallized intelligence, neither on average nor when allowing for gender heterogeneity (see Table 7).[48] This indicates that students affected by the reform have caught up with their non-affected counterparts in terms of mathematical competences. However, these estimates also reveal that the age-respective timing of instruction during adolescence does not influence skill formation in this crystallized domain; for two potential reasons: on the one hand, there may be neither a positive multiplier effect of earlier investment present nor an age effect benefiting older students' competence. On the other hand, both effects may be present, but offset each other. While it is not possible to disentangle these two mechanisms in this setting, the effects on fluid intelligence may give additional valuable insights.

Investigating the reform's effects on fluid measures of intelligence, the estimates show that there is no significant impact on processing speed; but reasoning ability, as measured by Raven's matrices test, is significantly lower for students affected by the reform. As fluid intelligence is assumed not to be directly affected by any type of investment, no positive

[47] Due to a resulting loss in the number of observations, I do not include school characteristics in this and the following estimations. Table A.15 shows however that, when further controlling for school characteristics, results are not altered.

[48] Even though a lack of statistical significance could be a consequence of the sample size, the 95% confidence intervals range between -0.14 and +0.14 (Males) and -0.09 and +0.12 (Females). Hence, in any case effect sizes would not be comparable to the impact of the increased instructional time.

Table 6: Average Effects of the Reform

	Outcome Variables: Cognitive Skills		
	Crystallized	Fluid	
	Mathematics	Speed	Reasoning
Reform	0.008	-0.048	-0.082*
	(0.045)	(0.066)	(0.047)
Female	-0.002	-0.002	0.004
	(0.056)	(0.055)	(0.050)
Migration background	-0.179***	-0.102	-0.167***
	(0.055)	(0.061)	(0.055)
Born January–June	0.063*	0.087*	-0.025
	(0.033)	(0.047)	(0.049)
High parental education	0.136**	-0.067	0.000
	(0.051)	(0.046)	(0.055)
Working-class father	-0.045	0.052	0.061
	(0.063)	(0.076)	(0.070)
Working mother	-0.029	-0.032	0.054
	(0.062)	(0.063)	(0.056)
Books at home	0.195***	0.010	0.013
	(0.048)	(0.055)	(0.046)
R^2	0.027	0.005	0.007
Observations	2125	2128	2128

Notes: NEPS:BW:3.0.0 wave 2011/2012. OLS regressions. Further, a constant is included. Standard errors, reported in parentheses, are clustered at the school level. * p<0.1, ** p<0.05, *** p<0.01.

multiplier effect of earlier instruction could be expected. Nonetheless, fluid intelligence *is* assumed and found to change with biological age where it is increasing during childhood and adolescence. The estimated decrease of around eight percent of a standard deviation, therefore, most likely stems from the age difference of one year, on average, between affected and non-affected students, but should not be related to the curriculum covered at any particular age. Still, these results on fluid intelligence can be taken into account when interpreting the zero effects on the crystallized dimension: Given the students' performance in the tasks to assess reasoning ability, age effects in cognitive skill formation seem still to be present in late adolescence benefiting older students. If this was true for all dimensions of cognition, age effects can be expected to also influence crystallized dimensions. Hence, the zero effect of the reform on mathematical ability among students of the same educational level may likely be the result of an interaction of this age effect

Table 7: Heterogeneous Effects of the Reform by Gender

	Outcome Variables: Cognitive Skills		
	Crystallized	Fluid	
	Mathematics	Speed	Reasoning
Reform	0.003	-0.066	-0.099
	(0.071)	(0.087)	(0.067)
Reform*Female	0.009	0.031	0.031
	(0.085)	(0.095)	(0.089)
R^2	0.027	0.005	0.007
Observations	2125	2128	2128

Notes: NEPS:BW:3.0.0 wave 2011/2012. OLS regressions. Individual characteristics controlled for include female, migration background, born January–June, high parental education, working-class father, working mother, and books at home. Further, a constant is included. Standard errors, reported in parentheses, are clustered at the school level. * p<0.1, ** p<0.05, *** p<0.01.

offset by a positive multiplier effect of earlier investment. The latter is underlined by the first analysis showing, based on SOEP data, that the earlier instruction could indeed be absorbed, at least by some students, leading to higher crystallized abilities among affected students at the age of seventeen. However, this male advantage in numerical skills at the age of seventeen did not translate into higher mathematical ability at the time of graduation.[49]

Although, on average, I find no significant effects on crystallized intelligence, the earlier timing of instruction may have impacted particular groups of students differently. If differential effects enlarge or reduce existing inequalities, they are of particular interest for policy makers seeking to decrease prevalent skill gaps. However, again, hardly any such differences exist with respect to demographic and socio-economic variables, the only exception hereto being socio-economic status when defined by the number of books at home where disadvantaged students face an improvement in mathematical ability.[50]

Analyzing whether the reform's effects differ by the characteristics of its implementation, reveals that the results are insensitive to the school's assessment of how smooth

[49]Note that numerical ability measured in SOEP is not directly comparable to mathematical ability measured in NEPS. Although both address the same or similar dimensions of crystallized intelligence, SOEP only tests basic numerical ability, independent of the educational curriculum covered. In contrast, the achievement test in NEPS explicitly asks for knowledge covered at this stage in high school including analysis, linear algebra, and statistics.

[50]Estimation results are available from the author upon request.

Table 8: Channels – Effects of the Reform on Leisure-time Activities and Paid Tutor Lessons

	Outcome Variables				
	Participation in Activity				Tutor
	Music	Sport	Reading	Computer	lessons
Reform	-0.010	0.029*	-0.034*	0.012	0.067**
	(0.021)	(0.015)	(0.017)	(0.010)	(0.026)
Reform	-0.009	0.004	-0.065**	0.015	0.080**
	(0.031)	(0.018)	(0.027)	(0.010)	(0.038)
Reform*Female	-0.001	0.043	0.054*	-0.006	-0.022
	(0.037)	(0.030)	(0.030)	(0.017)	(0.045)
Observations	2114	2096	2094	2086	2122

Notes: NEPS:BW:3.0.0 wave 2011/2012. OLS regressions. Individual characteristics controlled for include female, migration background, born January–June, high parental education, working-class father, working mother, and books at home. Further, a constant is included. Standard errors, reported in parentheses, are clustered at the school level. * p<0.1, ** p<0.05, *** p<0.01.

the implementation of the reform went.[51] In particular, whether certain aspects were regarded as strong stress factors in the transition or not, does not yield higher costs in terms of the school's students' cognitive abilities. In contrast, the perception of negative consequences of the reform did: in the schools where the reform was assessed to have a negative impact in general, students scored lower on the mathematics ability test. However, this may reflect the reverse pathway as low performing students may make the headmaster evaluate the reform negatively.

Potential Channels. In general, I find that crystallized intelligence is not affected by the earlier timing of instruction. One reason could be that generally the earlier instruction could simply have replaced cognitively stimulating activities at home. Unfortunately the data do not allow for the investigation of responses in leisure-time allocation to activities that are especially mathematically stimulating. Computer usage refers in particular to playing computer games and chatting, instead of programming, and is therefore not as mathematically challenging. Still, there is no reform effect on computer usage (see Table 8). In contrast, there are responses to the reform in terms of reading: it is especially male students who reduce reading in their leisure-time following the reform. This may explain

[51]Estimation results are available from the author upon request.

why the male advantage in crystallized intelligence obtained at the age of seventeen is offset until graduation as male students reduce their cognitive stimulating leisure-time activities compared to females.[52] For all students, I do, however, find an increase in the demand for paid tutor lessons following the reform by more than six percentage points. This suggests that some students may have had difficulties coping with the earlier instruction and the induced increased learning intensity. These adverse effects could also explain the zero finding, on average, of why a presumably positive multiplier effect of earlier instruction did not lead to higher mathematical ability. Contrary to the seventeen year-old respondents from the SOEP, in this sample of high school graduates male and female students seem to cope similarly with the higher requirements at school induced by the reform as the usage of paid tutor lessons increases statistically equally among both groups. The participation in sports during leisure-time seems to increase as well, at least on average, but the change amounts to only three percentage points.

7 Sensitivity Analyses

To confirm the positive effect of the increased instructional time, at least among male students, I conduct several robustness checks. Table A.9 presents methodological alterations: As mentioned in Section 5.1, wild cluster bootstrap to account for the small number of clusters yields even lower standard errors. Including a linear trend for each state to allow for state-specific developments over time, I also find that results are not altered. Further, to account for multiple hypotheses testing, I construct a summary index of cognitive ability following Anderson (2008). Using this measure, for which tests should be robust to overtesting as additional outcomes do not increase the probability of a false rejection, the estimation still reveals significant positive effects for male and zero effects for female students.[53]

[52]Reading may be even more important for the development of verbal skills. In future work, it would therefore be interesting to investigate the reform's effect on a further crystallized dimension of intelligence that entails more verbal aspects like the achievement test in English. Here one could suspect lower test scores related to the crowding out of reading during leisure-time among male students.

[53]The concern of multiple hypotheses testing arises when investigating several outcomes. Therefore, summary indexes weight each dimension to maximize the amount of information captured by the index. In addition, these tests may be more powerful and allow estimating a more general effect (Anderson,

To deal with the measurement of cognitive skills (see Table A.10), I first exclude all students aged eighteen or nineteen years from my sample to rule out potential age effects.[54] While the positive impact of instructional time on males' numerical skills is preserved, fluid intelligence scores, however, increase in this specification as well.[55] Second, I control for the month in which the assessment of cognitive skills took place as those interviewed later in the year have additional knowledge over students interviewed earlier, which has no impact though.

Next, I consider the composition of students to validate that the reform can indeed by regarded as a quasi-natural experiment (see Table A.11). First, I omit all individual characteristics in the specification. As this does not alter the results qualitatively, I conclude that they are not biased by the omission of these observables. Second, I add grade repeaters to my analysis which is important to consider in case there are systematic differences in repetition rates following the reform e.g. by gender. This seems however not to be the case, as the results are preserved.

Crucial for the interpretation of a causal effect in specification (2) is that there is no selectivity. As elaborated on in section 5.1, selectivity is unlikely but possible (see Table A.12). First, I consider selectivity within the sample occurring whenever students attend school in a different state where the reform was not yet implemented to avoid the educational change. As this entails high moving costs for the entire family, it is highly unlikely. Still, I consider two samples in which this is not possible or did not occur: (1) federal states that adopted the reform relatively late and neighbor states where the reform was already implemented[56] and (2) students who never moved from their place of childhood.[57] Both estimations show that the effects are completely preserved with the

2008).

[54]In all other specifications, I include birth year dummies to account for this.

[55]Important to note is, that it is unlikely that crystallized and fluid skills can be assessed completely separately in a test environment. Especially among male students of my sample, the correlation between numerical and figural skills amounts to 0.39. When looking at all other specifications, this effect is however not robust, which is in line with theory predicting that fluid skills should not be affected by educational changes other than through indirect effects via crystallized cognition.

[56]I define late-adopter states as all states where the first cohort affected by the reform graduated in 2012 or later. These include Baden-Wuerttemberg, Berlin, Brandenburg, Bremen, Hesse, North Rhine-Westphalia, and Schleswig-Holstein.

[57]These constitute 94.5% of my original sample, hence demonstrating very low regional mobility.

magnitude even rising to over 0.3 of a standard deviation increase in males' numerical skills. Second, I consider selectivity out of the sample, i.e. students choosing not to attend high school due to the reform, although this is highly unlikely as Huebener and Marcus (2015) find that high school entry rates and graduation rates are not affected by the reform. In some states, however, the university entrance qualification can still be obtained after a total of 13 years of schooling at comprehensive schools, which combine all three – academic, intermediate and lower – tracks of secondary school.[58] As comprehensive schools are not equally common across all federal states, I include only students from states where no or only few comprehensive schools exist to further rule out selection.[59] Qualitatively, effects are preserved in this sample, though coefficients drop in magnitude and loose significance, which may be attributable to the substantial loss in sample size.

Institutional Aspects are considered in Table A.13: First, even though there is a large discrepancy between former East and West Germany and their educational systems historically differed, the reform effects do not differ between these regions. Second, the exclusion of students from Saxony and Thuringia, where the reform was not introduced but an established twelve-year-schooling system continued, shows that the magnitude of the effect slightly decreases but that males' numerical skills still significantly improve and in particular benefit compared to females' skills. Third, I consider central exit examinations which existed since the 1990s or earlier in some states[60] and were introduced between 2005 and 2008 in most of the remaining states. Importantly, the introduction of these examinations did not coincide with the implementation of the high school reform, but took place earlier. Hence, the entire sample under analysis, both the treatment *and* control group, are subject to final exit examinations to obtain their *Abitur*. Moreover, central exit examinations are only relevant at the very end of high school and, hence, are unlikely to systematically affect seventeen year-olds at their current educational stage.

[58]Note that students from comprehensive schools are excluded from my sample. Still, the option could have affected selection into high school.

[59]I define states with no or only few comprehensive schools if the share of students attending comprehensive schools in this state is less than 10% between 2000 and 2013. These are Baden-Wuerttemberg, Bavaria, Lower Saxony, Mecklenburg-West Pommerania, Saxony, Saxony-Anhalt, and Thuringia (Autorengruppe Bildungsberichterstattung, 2012, 2014).

[60]Specifically Baden-Wuerttemberg, Bavaria, Mecklenburg-West Pomerania, Saarland, Saxony, Saxony-Anhalt and Thuringia.

Still, to rule out any interplay between the effects of both educational changes, I consider the subsample of states only where this tradition was long-standing existent. The direction of effects is preserved but coefficients decrease in magnitude. Also, statistical significance vanishes, possibly due to the substantial reduction in sample size. Fourth, I exclude students from the double cohort which might have been impacted differently, as the first cohort affected by the reform and the last one not affected were to merge into one class for the last two final years. Effect sizes are not altered; the drop in sample size may however again account for the loss in statistical significance. Fifth, I use dummies based on the academic years instead of calender years to account for the time dimension in the estimation of equation (2), as the academic year ranges from July to June, as defined by school-entry cut-off dates. The results are consistent with previous estimations.

Lastly, to verify the empirical strategy, I use the sample of students who follow secondary school tracks *other* than high school in a Placebo estimation where the reform indeed does not show effects (see Table A.14).

To confirm the validity of the findings on the timing of instruction, which exhibit zero effects in mathematics and slight decreases in fluid intelligence, I also conduct several sensitivity analyses. As this analysis is bound to one federal state, Baden-Wuerttemberg, no regional variation can be exploited to account for state-specific factors questioning complete external validity. Still, to prove robustness of the effects within Baden-Wuerttemberg, I focus on the composition of students (see Tables A.16 and A.17): First, I omit all individual characteristics from the specification which indicates that results are not biased by the omission of these observables as effects are not altered. Second, a weighted regression to allow generalization of the results with respect to high school graduates in Baden-Wuerttemberg, confirms the coefficient sizes, while statistical significance naturally drops (Chambers and Skinner, 2003). Third, I add grade repeaters to the sample, as Huebener and Marcus (2015) find no effects of the reform on repetition rates up to grade nine, but do find that rates doubled in the final years of high school. Results are not altered, as scores on mathematics are still unaffected while reasoning ability does decrease following the reform. Lastly, I add further waves to the sample to disentangle the effect potentially specific to the double-graduating cohort. For this I include one wave with students prior to the reform (2010/2011) and one with students post reform (2012/2013). The results show that the zero effect of mathematics is not altered, while the negative impact on reasoning ability drops. Hence, it is the students in the double-graduating cohort who specifically face a short-term decline in fluid scores. However, the results have to be taken cautiously, as the institutional-specific effect of being in the double-graduating cohort cannot be disentangled from time-specific effects[61].

8 Conclusion

As cognitive skills are important determinants of many economic and social outcomes, higher cognitive skills are often correlated with higher education. However, it is not only that individuals with higher cognitive abilities are likely to be better educated, but also that education improves cognitive skills. Most studies use changes in compulsory schooling laws as an exogenous variation to identify causal positive effects of an additional year of schooling. However, there is not much evidence on the underlying mechanisms in the economic literature.

This paper provides first evidence on disentangling two mechanisms through which education may improve cognitive skills in adolescence. I exploit a German high school reform carried out at the state-level between 2001 and 2007 as a quasi-natural experiment to estimate causal effects of this educational change on adolescents' cognitive abilities. Based on two separate analyses using SOEP and NEPS data, this study successfully disentangles the differential effects of instruction by focusing on quantity, on the one hand, and allocation with respect to age, on the other hand.

The improvement of crystallized intelligence through instructional time among seventeen-year old male students by up to 0.3 standard deviations is comparable to the effect sizes of one additional year of schooling in e.g. Scandinavian countries (cf. Brinch and Galloway, 2012; Carlsson et al., 2015; Falch and Massih, 2011). To the best of my knowledge,

[61] These could be e.g. transition rates into high school changing over time, inducing a different composition of students.

this is the first study pointing at important heterogeneous effects by gender indicating that educational quantity aggravates, instead of mitigates, gender skill differences by extending comparative advantages. This is of particular policy relevance for initiatives aiming at promoting female participation in the so-called STEM fields: the increasing gap in numerical skills may discourage women even further to enter mathematically oriented areas dominated by males.

In contrast, a positive multiplier-effect that could result from this skill acquisition at younger ages does not seem to outweigh potential biological age effects until graduation. The differential age-respective timing of educational instruction during adolescence does therefore not significantly alter cognitive skill development when comparing crystallized measures of competences of students affected by the reform and students not affected by the reform at the end of high school. As fluid intelligence is generally not assumed to change over the life cycle in response to factors other than age, no positive multiplier effect can be expected for the reform to increase fluid components of intelligence. The age gap therefore yields even lower scores for students affected by the reform compared to their non-affected counterparts. Lastly, these results can be drawn onto for the evaluation of the reform: they may justify the maintenance of the curriculum while shortening high school duration as students seem to absorb the higher load of subject matters taught. Apart from lower reasoning scores, which may be attributable to the age difference, the results suggest that high school graduates are just as equally well off before and after the reform in terms of acquired competences.

I conclude from these analyses that in the positive impact education has on cognitive skills the relevant factor is not (only) school duration but especially the amount of content taught. There is, however, important effect heterogeneity: With respect to gender, initial skill differences are further aggravated through an increased curriculum. For decision makers this opens up new possibilities to target cognitive ability, other than simply changing overall school duration when designing educational policies. However, differential effects need to be taken into consideration to avoid increasing inequality.

References

Amthauer, R., Brocke, B., Liepmann, D., and A., B. (2001). *Intelligenz-Struktur-Test 2000 R (I-S-T 2000 R)*. Göttingen: Hogrefe-Verlag.

Anderson, M. L. (2008). Multiple Inference and Gender Differences in the Effects of Early Intervention: A Reevaluation of the Abecedarian, Perry Preschool, and Early Training Projects. *Journal of the American Statistical Association*, 103(484):1481–1495.

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

Autorengruppe Bildungsberichterstattung (2010). *Bildung in Deutschland 2010*. W. Bertelsmann Verlag GmbH & Co, Bielefeld.

Autorengruppe Bildungsberichterstattung (2012). *Bildung in Deutschland 2012*. W. Bertelsmann Verlag GmbH & Co, Bielefeld.

Autorengruppe Bildungsberichterstattung (2014). *Bildung in Deutschland 2014*. W. Bertelsmann Verlag GmbH & Co, Bielefeld.

Baltes, P. B. (1987). Theoretical Propositions of Life-Span Developmental Psychology: On the Dynamics Between Growth and Decline. *Developmental Psychology*, 23(5):611–626.

Baltes, P. B. (1993). The Aging Mind: Potential and Limits. *The Gerontologist*, 33(5):580–594.

Banks, J. and Mazzonna, F. (2012). The Effect of Education on Old Age Cognitive Abilities: Evidence from a Regression Discontinuity Design. *Economic Journal*, 122(560):418–448.

Black, S. E., Devereux, P. J., and Salvanes, K. G. (2005). Why the Apple Doesn't Fall Far: Understanding Intergenerational Transmission of Human Capital. *American Economic Review*, 95(1):437–449.

Blossfeld, H.-P., Roßbach, H.-G., and von Maurice, J. (2011). Education as a lifelong process: The German National Educational Panel Study (NEPS). *Zeitschrift für Erziehungswissenschaft*, Special Issue 14.

Brinch, C. N. and Galloway, T. A. (2012). Schooling in adolescence raises IQ scores. *Proceedings of the National Academy of Sciences*, 109(2):425–430.

Cameron, A. C., Gelbach, J. B., and Miller, D. L. (2008). Bootstrap-Based Improvements for Inference with Clustered Errors. *The Review of Economics and Statistics*, 90(3):414–427.

Carlsson, M., Dahl, G. B., Öckert, B., and Rooth, D.-O. (2015). The Effect of Schooling on Cognitive Skills. *The Review of Economics and Statistics*, 97(3):533–547.

Cascio, E. U. and Lewis, E. G. (2006). Schooling and the Armed Forces Qualifying Test: Evidence from School-Entry Laws. *Journal of Human Resources*, 41(2):294–318.

Chambers, R. and Skinner, C. (2003). *Analysis of Survey Data*. Wiley Series in Survey Methodology. Wiley.

Clotfelter, C. T., Ladd, H. F., and Vigdor, J. L. (2015). The Aftermath of Accelerating Algebra: Evidence from District Policy Initiatives. *Journal of Human Resources*, 50(1):159–188.

Cortes, K. E., Goodman, J. S., and Nomi, T. (2015). Intensive Math Instruction and Educational Attainment: Long-Run Impacts of Double-Dose Algebra. *Journal of Human Resources*, 50(1):108–158.

Cunha, F. and Heckman, J. J. (2007). The Technology of Skill Formation. *American Economic Review*, 97(2):31–47.

Cunha, F. and Heckman, J. J. (2008). Formulating, Identifying and Estimating the Technology of Cognitive and Noncognitive Skill Formation. *Journal of Human Resources*, 43(4):738–782.

Dahmann, S. and Anger, S. (2014). The Impact of Education on Personality: Evidence from a German High School Reform. IZA Discussion Papers 8139, Institute for the Study of Labor (IZA).

Doolittle, A. E. (1989). Gender Differences in Performance on Mathematics Achievement Items. *Applied Measurement in Education*, 2(2):161–177.

Ellison, G. and Swanson, A. (2012). Heterogeneity in High Math Achievement Across Schools: Evidence from the American Mathematics Competitions. NBER Working Papers 18277, National Bureau of Economic Research, Inc.

Falch, T. and Massih, S. S. (2011). The Effect Of Education On Cognitive Ability. *Economic Inquiry*, 49(3):838–856.

Hanushek, E. A. and Woessmann, L. (2008). The Role of Cognitive Skills in Economic Development. *Journal of Economic Literature*, 46(3):607–68.

Heckman, J. J., Stixrud, J., and Urzua, S. (2006). The Effects of Cognitive and Noncognitive Abilities on Labor Market Outcomes and Social Behavior. *Journal of Labor Economics*, 24(3):411–482.

Hedges, L. V. and Nowell, A. (1995). Sex Differences in Mental Test Scores, Variability, and Numbers of High-Scoring Individuals. *Science*, 269(5220):41–45.

Heineck, G. and Anger, S. (2010). The returns to cognitive abilities and personality traits in Germany. *Labour Economics*, 17(3):535–546.

Huebener, M. and Marcus, J. (2015). Moving up a Gear: The Impact of Compressing Instructional Time into Fewer Years of Schooling. Discussion Papers of DIW Berlin 1450, DIW Berlin, German Institute for Economic Research.

Jacobs, J. E., Lanza, S., Osgood, D. W., Eccles, J. S., and Wigfield, A. (2002). Changes in Childrens Self-Competence and Values: Gender and Domain Differences across Grades One through Twelve. *Child Development*, 73(2):509–527.

Kamhöfer, D. A. and Schmitz, H. (2015). Reanalyzing Zero Returns to Education in Germany. *Journal of Applied Econometrics, forthcoming.*

Krashinsky, H. (2014). How Would One Extra Year of High School Affect Academic Performance in University? Evidence from an Educational Policy Change. *Canadian Journal of Economics*, 47(1):70–97.

Lang, F., Weiss, D., Stocker, A., and von Rosenbladt, B. (2007). Assessing Cognitive Capacities in Computer-Assisted Survey Research: Two Ultra-Short Tests of Intellectual Ability in the German Socio-Economic Panel (SOEP). *Schmollers Jahrbuch: Journal of Applied Social Science Studies / Zeitschrift für Wirtschafts- und Sozialwissenschaften*, 127(1):183–192.

Lynn, R. (1994). Sex differences in intelligence and brain size: A paradox resolved. *Personality and Individual Differences*, 17(2):257–271.

Machin, S. and McNally, S. (2008). The literacy hour. *Journal of Public Economics*, 92(5-6):1441–1462.

Malecki, A., Schneider, C., Vogel, S., and Wolters, M. (2014). *Schulen auf einen Blick*. Statistisches Bundesamt, Wiesbaden.

Marcotte, D. E. (2007). Schooling and test scores: A mother-natural experiment. *Economics of Education Review*, 26(5):629–640.

OECD (2014). Promoting Female Participation in STEM. *What it means for higher education: quarterly briefs 2014*. Available on http://www.oecd.org/edu/imhe/Promoting-female-participation-in-STEM.pdf. Higher Education Programme, Organisation for Economic Co-operation and Development (OECD), Paris.

Richter, D., Metzing, M., Weinhardt, M., and Schupp, J. (2013). SOEP scales manual. SOEP Survey Papers 138, DIW Berlin, German Institute for Economic Research.

Schneeweis, N., Skirbekk, V., and Winter-Ebmer, R. (2014). Does Education Improve Cognitive Performance Four Decades After School Completion? *Demography*, 51(2):619–643.

Schupp, J. and Herrmann, S. (2009). Kognitionspotenziale Jugendlicher: Ergänzung zum Jugendfragebogen der Längsschnittstudie Sozio-oekonomisches Panel (SOEP); Methodenbericht und Variablendokumentation. Data Documentation 43, DIW Berlin, German Institute for Economic Research.

Spinath, B., Freudenthaler, H. H., and Neubauer, A. C. (2010). Domain-specific school achievement in boys and girls as predicted by intelligence, personality and motivation. *Personality and Individual Differences*, 48(4):481–486.

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

Tinklin, T. (2003). Gender Differences and High Attainment. *British Educational Research Journal*, 29(3):307–325.

Wagner, G. G., Frick, J. R., and Schupp, J. (2007). The German Socio-Economic Panel Study (SOEP) – Scope, Evolution and Enhancements. *Schmollers Jahrbuch: Journal of Applied Social Science Studies / Zeitschrift für Wirtschafts- und Sozialwissenschaften*, 127(1):139–169.

Warm, T. (1989). Weighted likelihood estimation of ability in item response theory. *Psychometrika*, 54(3):427–450.

Wigfield, A., Battle, A., Keller, L. B., and Eccles, J. S. (2002). Sex differences in motivation, self-concept, career aspiration, and career choice: Implications for cognitive development. In De Lisi, R. and De Lisi, A. M., editors, *Biology, society, and behavior: The development of sex differences in cognition*, Advances in Applied Developmental Psychology, vol. 21., pages 93–124. Ablex Publishing, Westport, CT, US.

A Appendix

A.1 Variables

Table A.1: Description of Variables in SOEP

Variable	Description
Verbal	Standardized measure for verbal skills
Numerical	Standardized measure for numerical skills
Figural	Standardized measure for figural skills
Age	Age (in years)
Female	Dummy for female
Migration background	Dummy for student with a migration background
Born January–June	Dummy for being born between January and June
Low-performing student	Dummy for having received a recommendation for a different type of secondary school, i.e. *other than high school*, after grade four
Rural area	Dummy for having lived most of the childhood until age 15 in rural area
High parental education	Dummy for at least one of an individual's parents having an upper secondary school degree or higher
Working-class father	Dummy for father having blue-collar occupation when student is aged 15, reference category encompasses *all* others
Working mother	Dummy for working mother (both full-time and part-time) when student is aged 10
Single parent	Dummy for not having lived with *both* parents for the entire time up to age 15
Paid tutor lessons	Dummy for attending paid tutor lessons (additional to regular school attendance)
Parents: Interest	Dummy for parents showing interest in school performance *'quite a lot'* or *'very much'*
Parents: Homework	Dummy for at least one parent helping with homework and studying
Parents: Problems	Dummy for having disagreements over studies with at least one parent
Music	Dummy for being musically active
Sport	Dummy for doing sports
Reading	Dummy for reading in leisure-time at least once a week
Tech. work	Dummy for doing technical work or programming in leisure-time at least once a week

Notes: SOEPv30 waves 2006 to 2013.

Table A.2: Description of Variables in NEPS

Variable	Description
Individual level	
Mathematics	Standardized WLE estimate of mathematical ability
Speed	Standardized measure for processing speed
Reasoning	Standardized measure for reasoning ability
Female	Dummy for female
Migration background	Dummy for at least one of the student's parents born abroad
Born January–June	Dummy for being born between January and June
High parental education	Dummy for at least one of an individual's parents having an upper secondary school degree or higher
Working-class father	Dummy for father currently having blue-collar occupation, reference category encompasses *all* others
Working mother	Dummy for mother currently working (both full-time and part-time)
Books at home	Dummy for having a book shelve of at least average size at home
Music	Dummy for participating in the orchestra or church groups at least once a week
Sport	Dummy for doing sports at least once a week
Reading	Dummy for reading in leisure-time at least once a week
Computer	Dummy for playing computer games, chatting etc. at least once a week
Tutor lessons	Dummy for attending paid tutor lessons (additional to regular school attendance)
School level	
Mig. backgr. (Cohort)	Dummy for share of students with a migration background in cohort >10%
Mig. backgr. (School)	Dummy for share of students with a migration background in school >10%
Mig. backgr. (Teachers)	Dummy for teachers with a migration background at school
School size	Dummy for above median total number of students at school
Negative consequences	Dummy for negative effects of reform visible
Stress: Organization	Dummy for organization as strong stress factor due to implementation of reform
Stress: Course scheme	Dummy for course scheme as strong stress factor due to implementation of reform
Stress: Counseling	Dummy for providing counseling as strong stress factor due to implementation of reform
Stress: Room	Dummy for room situation as strong stress factor due to implementation of reform
Stress: Material	Dummy for availability of teaching materials as strong stress factor due to implementation of reform

Notes: NEPS:BW:3.0.0 wave 2011/2012. School characteristics are self-reported by the headmaster.

A.2 Summary Statistics

Table A.3: Summary Statistics of (non-standardized) Scores on Cognitive Skills Tests in SOEP

	Observations	Mean	Std. Dev.	Minimum	Maximum
Males					
Verbal	335	11.236	3.235	3	20
Numerical	335	15.791	3.782	3	20
Figural	335	11.191	3.131	3	18
Females					
Verbal	388	10.389	3.176	1	18
Numerical	388	14.090	4.138	3	20
Figural	388	11.332	2.858	1	18

Notes: SOEPv30 waves 2006 to 2013, sample: adolescents aged 17 to 19 attending high school.

Table A.4: Gender Differences in (non-standardized) Scores on Cognitive Skills Tests in SOEP

	Mean		Equality of Means	
	Male	Female	Difference	t-stat
Control Group				
Verbal	11.303	10.458	0.844	2.702
Numerical	15.636	14.346	1.290	3.284
Figural	11.221	11.396	-0.175	-0.598
Observations	195	240		
Treatment Group				
Verbal	11.143	10.277	0.865	2.327
Numerical	16.007	13.676	2.331	5.183
Figural	11.150	11.230	-0.080	-0.232
Observations	140	148		

Notes: SOEPv30 waves 2006 to 2013, sample: adolescents aged 17 to 19 attending high school.

Table A.5: Summary Statistics of Individual Characteristics in SOEP

	Mean		Equality of Means
	Control	Treatment	t-stat
Age	17.474	17.083	8.236
Female	0.552	0.514	0.998
Migration background	0.161	0.167	-0.204
Born January–June	0.552	0.458	2.467
Low-performing student	0.168	0.194	-0.915
Rural area	0.278	0.326	-1.389
High parental education	0.618	0.538	2.147
Working-class father	0.168	0.219	-1.717
Working mother	0.736	0.806	-2.170
Single parent	0.200	0.188	0.415
Observations	435	288	

Notes: SOEPv30 waves 2006 to 2013, sample: adolescents aged 17 to 19 attending high school. The age differs by construction of the sample, as only in wave 2006 eighteen and nineteen year-olds were included in the adolescent sample as well. As at this point in time, in most states the reform was not implemented yet, the age is higher among the control group.

Table A.6: Summary Statistics of (non-standardized) Scores on Cognitive Skills Tests in NEPS

	Observations	Mean	Std. Dev.	Minimum	Maximum
Males					
Mathematics	932	0.383	1.091	-2.689	3.712
Speed	933	63.174	12.576	0	93
Reasoning	933	10.937	1.136	5	12
Females					
Mathematics	1193	-0.265	1.043	-5.027	3.712
Speed	1195	66.300	10.548	0	93
Reasoning	1195	10.643	1.321	1	12

Notes: NEPS:BW:3.0.0 wave 2011/2012, sample: high school students in final grade born between 1991 and 1995.

Table A.7: Gender Differences in (non-standardized) Scores on Cognitive Skills Tests in NEPS

	Mean		Equality of Means	
	Male	Female	Difference	t-stat
Control Group				
Mathematics	0.388	-0.266	0.654	9.649
Speed	63.608	66.492	-2.884	-4.077
Reasoning	10.998	10.687	0.311	4.021
Observations	444 (443)	571 (570)		
Treatment Group				
Mathematics	0.378	-0.263	0.641	10.018
Speed	62.779	66.125	-3.346	-4.715
Reasoning	10.881	10.603	0.279	3.661
Observations	489	624 (623)		

Notes: NEPS:BW:3.0.0 wave 2011/2012, sample: high school students in final grade born between 1991 and 1995. For Mathematics there are, in total, three observations less available than for the fluid skill measures Speed and Reasoning, resulting in minimally differing number of observations between the cognitive skill dimensions in some subgroups.

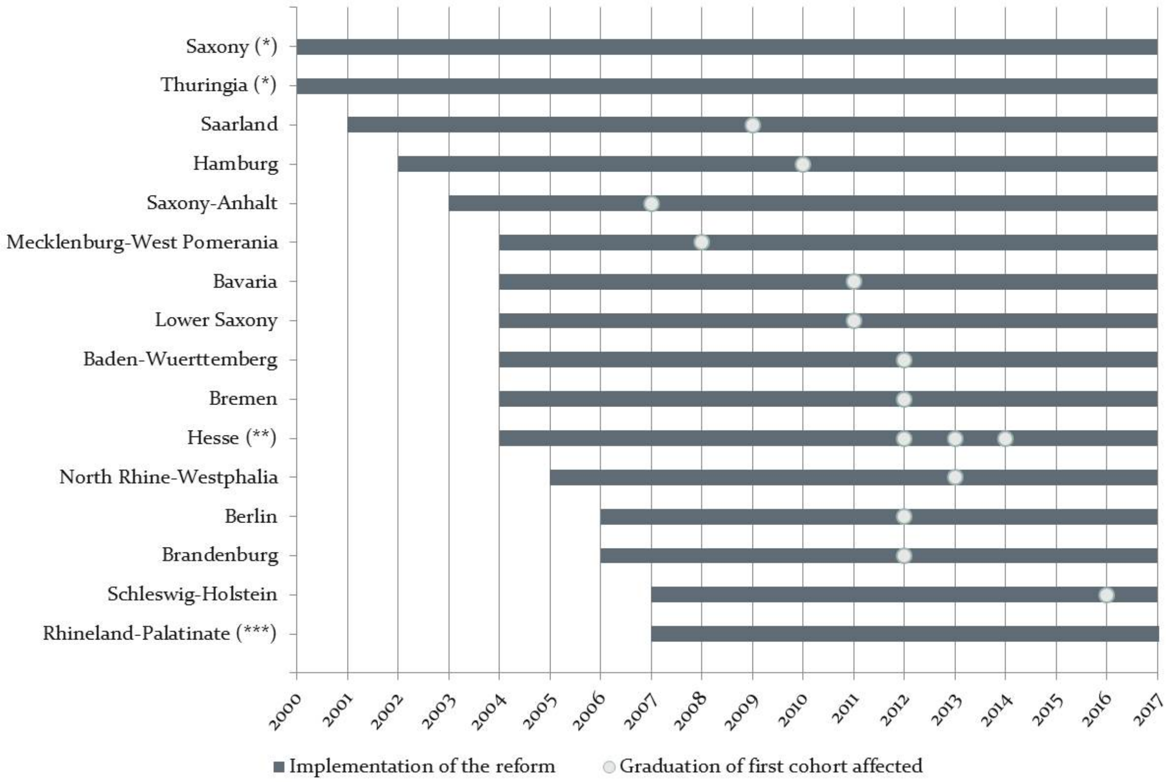
Table A.8: Summary Statistics of Individual Characteristics in NEPS

	Mean		Equality of Means
	Control	Treatment	t-stat
Age	19.431	18.486	40.530
Female	0.563	0.561	0.089
Migration background	0.207	0.202	0.271
Born January–June	0.472	0.438	1.591
High parental education	0.634	0.614	0.990
Working-class father	0.170	0.186	-0.935
Working mother	0.864	0.853	0.752
Books at home	0.665	0.635	1.439
Observations	1015	1113	

Notes: NEPS:BW:3.0.0 wave 2011/2012, sample: high school students in final grade born between 1991 and 1995.

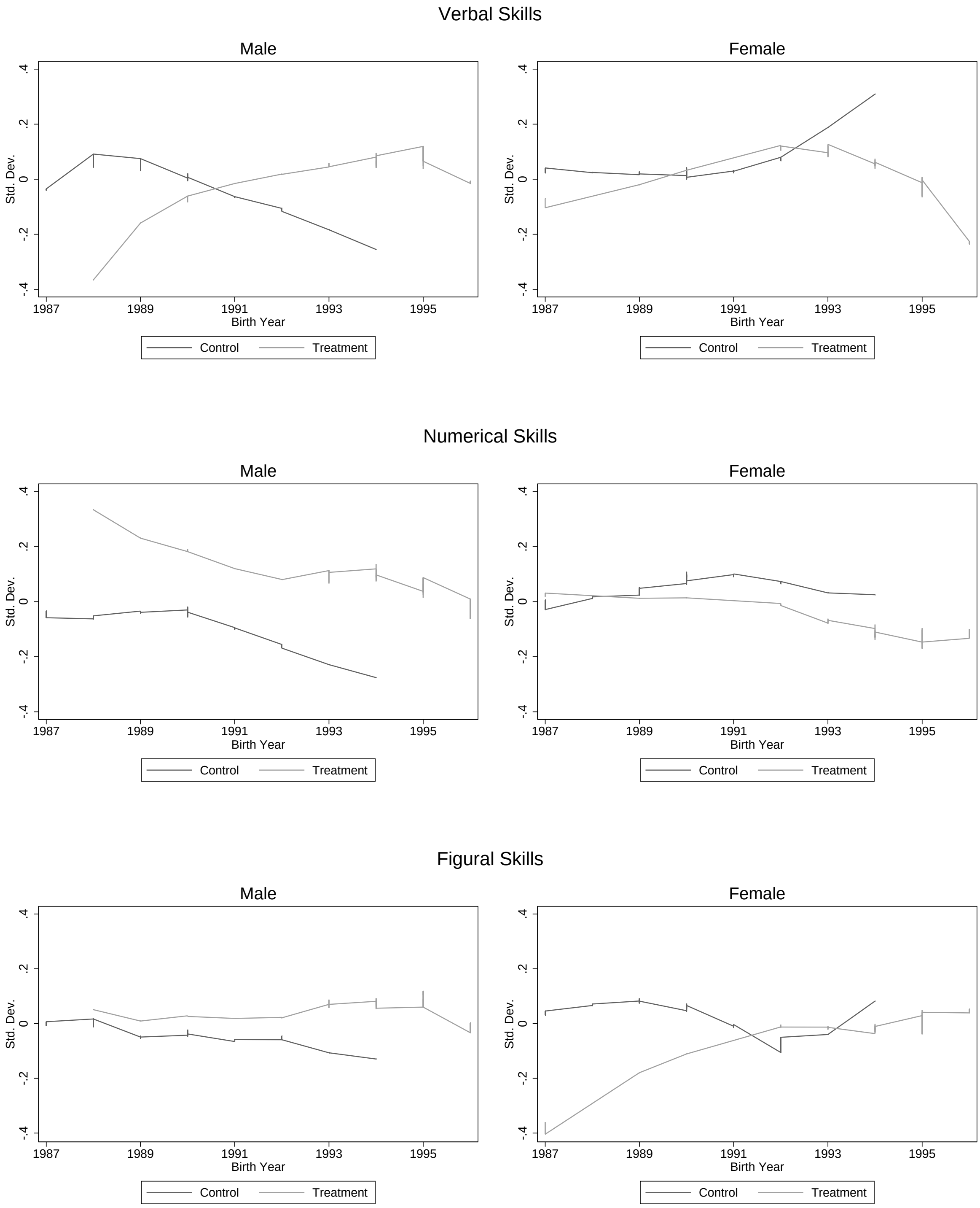
A.3 Figures

Figure A.1: Introduction of the Reform by Federal State



(*) Saxony and Thuringia established a 12-year school system before the 1990s.
(**) Gradual introduction for students entering high school in 2004/05 (10% of all schools), 2005/06 (60%), and 2006/07 (30%).
(***) In Rhineland-Palatinate the reform has only been introduced in selected schools.
Source: Autorengruppe Bildungsberichterstattung (2010)

Figure A.2: Instructional Time – Cognitive Skill Measures (in standard deviations) of adolescents in SOEP over Cohorts by Treatment Status



Notes: SOEPv30 waves 2006 to 2013. Lowess Plots of age-free measures in standard deviations of cognitive skills. These are obtained as residuals from a regression of the cognitive skill measure on age and age squared, to account for potential age effects. Birth Year-Treatment Status combinations with less than 5 observations are excluded.

A.4 Estimation Results on Instructional Time

Table A.9: Sensitivity Analyses – Methodology

	Outcome Variables: Cognitive Skills		
	Crystallized		Fluid
	Verbal	Numerical	Figural
Wild cluster bootstrapped standard errors[1]			
Reform	0.097	0.297**	0.144
	(0.097)	(0.116)	(0.112)
Reform*Female	-0.053	-0.296***	-0.103
	(0.151)	(0.088)	(0.097)
Observations	723	723	723
Including linear state time trends[2]			
Reform	0.127	0.340**	0.190
	(0.125)	(0.115)	(0.126)
Reform*Female	-0.041	-0.286**	-0.083
	(0.164)	(0.099)	(0.105)
Observations	723	723	723
Accounting for the problem of multiple hypotheses testing			
	Outcome Variable: Overall Cognitive Skill Measure[3]		
Reform	0.182*		
	(0.087)		
Reform*Female	-0.154**		
	(0.067)		
Observations	723		

Notes: SOEPv30 waves 2006 to 2013. OLS regressions. Further, a maximum set of state dummies, year of birth dummies, dummies for the different SOEP subsamples, and a constant are included. Individual characteristics controlled for include female, migration background, born January–June, low-performing student, rural area, high parental education, working-class father, working mother, and single parent. Standard errors, reported in parentheses, are clustered at the state level. * p<0.1, ** p<0.05, *** p<0.01.
[1]The reported standard errors are wild cluster bootstrapped following Cameron et al. (2008).
[2]This estimation further includes a linear time trend for each state.
[3]Overall cognitive skill measure obtained following Anderson (2008) to avoid the problem of multiple hypotheses testing.

Table A.10: Sensitivity Analyses – Measurement

	Outcome Variables: Cognitive Skills		
	Crystallized		Fluid
	Verbal	Numerical	Figural
Subsample: Age 17			
Reform	0.220	0.255**	0.235*
	(0.148)	(0.110)	(0.117)
Reform*Female	-0.097	-0.273**	-0.130
	(0.172)	(0.110)	(0.112)
Observations	568	568	568
Controlling for month of interview[1]			
Reform	0.097	0.300**	0.131
	(0.096)	(0.129)	(0.124)
Reform*Female	-0.049	-0.308**	-0.115
	(0.159)	(0.106)	(0.110)
Observations	721	721	721

Notes: SOEPv30 waves 2006 to 2013. OLS regressions. Further, a maximum set of state dummies, year of birth dummies, dummies for the different SOEP subsamples, and a constant are included. Individual characteristics controlled for include female, migration background, born January–June, low-performing student, rural area, high parental education, working-class father, working mother, and single parent. Standard errors, reported in parentheses, are clustered at the state level. * p<0.1, ** p<0.05, *** p<0.01.
[1]The estimation includes a dummy for each month of interview.

Table A.11: Sensitivity Analyses – Student Composition

	Outcome Variables: Cognitive Skills		
	Crystallized		Fluid
	Verbal	Numerical	Figural
No individual characteristics included			
Reform	0.085	0.252*	0.105
	(0.114)	(0.119)	(0.113)
Reform*Female	-0.039	-0.151**	-0.059
	(0.128)	(0.060)	(0.074)
Observations	723	723	723
Inclusion of grade repeaters[1]			
Reform	0.117	0.324**	0.169
	(0.095)	(0.125)	(0.110)
Reform*Female	-0.088	-0.316*	-0.138
	(0.145)	(0.148)	(0.096)
Observations	778	778	778

Notes: SOEPv30 waves 2006 to 2013. OLS regressions. Further, a maximum set of state dummies, year of birth dummies, dummies for the different SOEP subsamples, and a constant are included. Standard errors, reported in parentheses, are clustered at the state level. * p<0.1, ** p<0.05, *** p<0.01.

[1]This estimation further includes female, migration background, born January–June, low-performing student, rural area, high parental education, working-class father, working mother, single parent, and having repeated a grade as individual control variables.

Table A.12: Sensitivity Analyses – Selectivity

	Outcome Variables: Cognitive Skills		
	Crystallized		Fluid
	Verbal	Numerical	Figural
Subsample: Late adopter states[1]			
Reform	0.395	0.304***	0.061
	(0.320)	(0.079)	(0.321)
Reform*Female	0.014	-0.372	-0.052
	(0.388)	(0.193)	(0.137)
Observations	412	412	412
Subsample: Students who never moved from place of childhood			
Reform	0.141	0.338**	0.137
	(0.113)	(0.118)	(0.114)
Reform*Female	-0.069	-0.343***	-0.154
	(0.139)	(0.097)	(0.111)
Observations	683	683	683
Subsample: States with no or few comprehensive schools			
Reform	-0.016	0.218	0.053
	(0.120)	(0.186)	(0.193)
Reform*Female	-0.079	-0.090	0.094
	(0.195)	(0.154)	(0.094)
Observations	395	395	395

Notes: SOEPv30 waves 2006 to 2013. OLS regressions. Further, a maximum set of state dummies, year of birth dummies, dummies for the different SOEP subsamples, and a constant are included. Individual characteristics controlled for include female, migration background, born January–June, low-performing student, rural area, high parental education, working-class father, working mother, and single parent. Standard errors, reported in parentheses, are clustered at the state level. * p<0.1, ** p<0.05, *** p<0.01.
[1]These are states where the first students affected by the reform graduate in 2012 or later, that is, Baden-Wuerttemberg, Bremen, Hesse, North Rhine-Westphalia, Berlin, Schleswig-Holstein, and Brandenburg.

Table A.13: Sensitivity Analyses – Institutional Factors

	Outcome Variables: Cognitive Skills		
	Crystallized		Fluid
	Verbal	Numerical	Figural
Heterogeneous Effects for former East and West Germany			
Reform	0.037	0.133	0.076
	(0.096)	(0.138)	(0.109)
Reform*East	0.141	0.046	0.064
	(0.196)	(0.242)	(0.144)
Observations	723	723	723
Subsample: Exclusion of Saxony and Thuringia			
Reform	0.104	0.180*	0.146
	(0.133)	(0.095)	(0.119)
Reform*Female	0.042	-0.343***	-0.064
	(0.156)	(0.097)	(0.113)
Observations	663	663	663
Subsample: States with long-time standing central exit examinations			
Reform	-0.078	0.206	-0.020
	(0.168)	(0.219)	(0.224)
Reform*Female	-0.096	-0.035	0.103
	(0.230)	(0.208)	(0.126)
Observations	323	323	323
Subsample: Exclusion of double graduating cohort			
Reform	-0.126	0.251	-0.124
	(0.129)	(0.216)	(0.222)
Reform*Female	-0.054	-0.213	0.103
	(0.132)	(0.147)	(0.179)
Observations	582	582	582
Year dummies for academic year (instead of calender year)			
Reform	0.120	0.271*	0.102
	(0.112)	(0.130)	(0.105)
Reform*Female	-0.038	-0.292***	-0.076
	(0.145)	(0.088)	(0.114)
Observations	723	723	723

Notes: SOEPv30 waves 2006 to 2013. OLS regressions. Further, a maximum set of state dummies, year of birth dummies, dummies for the different SOEP subsamples, and a constant are included. Individual characteristics controlled for include female, migration background, born January–June, low-performing student, rural area, high parental education, working-class father, working mother, and single parent. Standard errors, reported in parentheses, are clustered at the state level. * $p<0.1$, ** $p<0.05$, *** $p<0.01$.

Table A.14: Placebo Estimation – Average Effects of the Reform on Students from Secondary School Tracks other than High School and Comprehensive School

	Outcome Variables: Cognitive Skills		
	Crystallized		Fluid
	Verbal	Numerical	Figural
Reform	-0.062	-0.006	-0.038
	(0.096)	(0.138)	(0.104)
Reform*Female	-0.159	-0.057	0.163
	(0.132)	(0.093)	(0.095)
Female	0.110	0.051	-0.076
	(0.069)	(0.057)	(0.095)
Migration background	-0.203***	-0.189*	-0.282***
	(0.053)	(0.094)	(0.078)
Born January–June	0.067	0.066	0.046
	(0.071)	(0.071)	(0.069)
Low-performing student	-0.054	0.044	0.081
	(0.060)	(0.056)	(0.059)
Rural area	0.120	0.115	0.058
	(0.090)	(0.077)	(0.071)
High parental education	0.371***	0.130	0.262***
	(0.111)	(0.089)	(0.075)
Working-class father	-0.238***	-0.057	-0.088
	(0.060)	(0.062)	(0.058)
Working mother	0.069	0.165**	0.149**
	(0.054)	(0.062)	(0.056)
Single parent	-0.021	-0.059	-0.083
	(0.067)	(0.049)	(0.065)
Repeated grade	-0.198*	-0.119*	-0.194*
	(0.098)	(0.059)	(0.107)
R^2	0.139	0.093	0.107
Observations	964	964	964

Notes: SOEPv30 waves 2006 to 2013. OLS regressions. Further, a maximum set of state dummies, year of birth dummies, dummies for the different SOEP subsamples, and a constant are included. Standard errors, reported in parentheses, are clustered at the state level. * p<0.1, ** p<0.05, *** p<0.01.

A.5 Estimation Results on Timing of Instruction

Table A.15: Effects of the Reform (controlling for school characteristics)

	Outcome Variables: Cognitive Skills		
	Crystallized	Fluid	
	Mathematics	Speed	Reasoning
Reform	-0.036	-0.079	-0.110**
	(0.048)	(0.072)	(0.050)
Female	-0.005	0.005	0.042
	(0.060)	(0.058)	(0.054)
Migration background	-0.137**	-0.059	-0.181***
	(0.058)	(0.066)	(0.066)
Born January–June	0.066*	0.064	-0.029
	(0.036)	(0.050)	(0.054)
High parental education	0.103*	-0.056	-0.014
	(0.054)	(0.043)	(0.060)
Working-class father	-0.072	0.032	0.064
	(0.063)	(0.086)	(0.078)
Working mother	-0.025	-0.064	0.081
	(0.071)	(0.066)	(0.061)
Books at home	0.199***	0.022	0.013
	(0.052)	(0.058)	(0.052)
Mig. backgr. (cohort)	-0.211***	-0.162	-0.173**
	(0.072)	(0.100)	(0.072)
Mig. backgr. (school)	-0.070	-0.223*	0.097
	(0.093)	(0.123)	(0.091)
Mig. backgr. (teacher)	0.017	0.130	-0.086
	(0.100)	(0.116)	(0.078)
School size	-0.046	-0.233**	0.085
	(0.083)	(0.110)	(0.065)
Negative consequences	0.233***	0.099	0.071
	(0.073)	(0.129)	(0.088)
Stress: Course scheme	-0.159	0.125	0.065
	(0.096)	(0.127)	(0.068)
Stress: Counseling	0.187**	0.222*	-0.084
	(0.090)	(0.125)	(0.070)
Stress: Room	0.104	0.048	-0.111*
	(0.067)	(0.091)	(0.062)
Stress: Material	-0.106	0.082	-0.015
	(0.088)	(0.100)	(0.080)
R^2	0.050	0.040	0.022
N	1793	1796	1796

Notes: NEPS:BW:3.0.0 wave 2011/2012. OLS regressions. Further, a constant is included. Standard errors, reported in parentheses, are clustered at the school level. *p<0.1, **p<0.05, ***p<0.01.

Table A.16: Sensitivity Analyses – Student Composition I

	Outcome Variables: Cognitive Skills		
	Crystallized	Fluid	
	Mathematics	Speed	Reasoning
No individual characteristics included			
Reform	-0.002	-0.048	-0.081*
	(0.045)	(0.066)	(0.047)
Reform	-0.009	-0.066	-0.102
	(0.073)	(0.088)	(0.068)
Reform*Female	0.012	0.031	0.039
	(0.087)	(0.093)	(0.090)
Observations	2125	2128	2128
Weighted regression[1]			
Reform	0.025	-0.013	-0.071
	(0.048)	(0.070)	(0.048)
Reform	0.039	-0.011	-0.077
	(0.077)	(0.096)	(0.072)
Reform*Female	-0.024	-0.004	0.012
	(0.098)	(0.103)	(0.100)
Observations	2125	2128	2128

Notes: NEPS:BW:3.0.0 wave 2011/2012. OLS regressions. Female is controlled for and further a constant is included. Standard errors, reported in parentheses, are clustered at the school level. * p<0.1, ** p<0.05, *** p<0.01.

[1]This estimation further controls for individual characteristics including migration background, born January–June, high parental education, working-class father, working mother, and books at home.

Table A.17: Sensitivity Analyses – Student Composition II

	Outcome Variables: Cognitive Skills		
	Crystallized	Fluid	
	Mathematics	Speed	Reasoning
Inclusion of grade repeaters			
Reform	0.035	-0.049	-0.082*
	(0.043)	(0.065)	(0.046)
Reform	0.043	-0.061	-0.100
	(0.069)	(0.088)	(0.069)
Reform*Female	-0.014	0.022	0.032
	(0.085)	(0.093)	(0.082)
Observations	2235	2238	2238
Additional inclusion of waves 2010/2011 and 2012/2013[1]			
Reform	0.046	-0.024	-0.033
	(0.046)	(0.064)	(0.039)
Reform	-0.011	-0.013	0.027
	(0.062)	(0.072)	(0.059)
Reform*Female	0.101	-0.019	-0.106
	(0.063)	(0.068)	(0.096)
Observations	4230	4235	4235

Notes: NEPS:BW:3.0.0 wave 2011/2012. OLS regressions. Individual characteristics controlled for include female, migration background, born January–June, high parental education, working-class father, working mother, and books at home. Further, a constant is included. Standard errors, reported in parentheses, are clustered at the school level. * $p<0.1$, ** $p<0.05$, *** $p<0.01$.
[1]Next to wave 2011/2012, these estimations include data from waves 2010/2011 and 2012/2013. Further, a maximum set of dummies indicating the wave are included in the specification.